

Retractions' realities

The papers are published, the news is out, the kudos is gathered, the gasps of admiration over. But the results are wrong. What happens next?

It is regrettable but inevitable that the scientific record contains errors. Many are relatively trivial. Others are more substantive, but ultimately reflect the progress of science itself. Their publication is representative of the state of knowledge at the time, so they do not require formal correction or withdrawal from the literature.

A much rarer beast is a paper established to be fundamentally flawed around the time of publication, when the results in question are still very much common parlance. For other researchers, the consequences, in terms of cost and time, of such errors propagating in the literature are very real. It is therefore incumbent on the authors of questionable work and the editors of the journals who host it to endeavour to set the scientific record straight when a mistake comes to light. And if the problem is sufficiently serious, calling into question the main findings of a piece of work, the most effective route of correction is through retraction of the paper itself.

Nature this week finds itself in the unenviable (and unprecedented) position of formally retracting seven papers (see page 92). All share the same first author, Jan Hendrik Schön (see *Nature* **419**, 417; 2002); in fact, this represents the entire body of work published by Schön in this journal.

The case against five of the papers is clear cut: they were subject to scrutiny by an investigating committee and were judged to show clear evidence of scientific misconduct. A statement of intent to retract these papers was rapidly forthcoming from most of the authors, and *Nature* published due warnings. But, as will be clear from the published retractions, the final wording is manifestly a compromise between the various parties involved.

The authors clearly wish to go on record as retaining some confidence in the validity of the science behind the misreported experimental findings; indeed, the scientific plausibility of what had been claimed may in part reflect the ease with which the original publications survived the rigours of peer review. And in one case, no clear agreement could be obtained on a common wording, not least

because two groups of co-authors refused point-blank to co-sign a statement with Schön. Such disputes between co-authors are not unusual, typically leading to considerable delays. But what ultimately matters is that these papers are now retracted.

But what of the other two? These did not fall within the remit of the investigation — which, in order to be prompt, restricted its scope to just a subset of Schön's vast output — nor have they subsequently been found to be formally lacking. So why are they nevertheless being retracted? Given the wide-ranging nature of the committee's findings concerning Schön's research practices, all of the authors on these two non-investigated papers were invited by this journal to consider their position with regard to this work. *Nature* did not insist on retractions, yet the authors (with the exception of Schön) were no longer prepared to support the conclusions of these papers and decided that they too should be withdrawn. Whereas Schön remains insistent that his results should stand unless faced with hard evidence to the contrary (a frustrating position that he maintained throughout the investigation), his former colleagues and collaborators have now assumed in full their scientific responsibilities as co-authors on the papers in question.

Clarity is critical to the published scientific literature. The sad story of Schön is thankfully an exception, rather than the rule: the vast majority of errors represent honest mistakes, and a journal needs to distinguish clearly between these. *Nature's* policy guidelines at www.nature.com/nature/submit/policies guide authors through this and other minefields of scientific conduct. Journals with online versions of their papers can further illuminate this darkness by specifying and highlighting mistakes made, whether by authors or by the journals themselves. Linking forwards and backwards between the paper and its subsequent correction ensures that information is public, transparent and accurate. The integrity of the scientific record is thereby maintained, and the process of moving forwards in discovery is all the better for it. ■

Yes, we have no energy policy

A plan to generate electricity from coal without emitting greenhouse gas is as threadbare as the rest of Bush's energy policy.

President George W. Bush has been making several attempts of late to demonstrate that he is serious about tackling the separate but related problems of global climate change and the United States' dependency on imported oil.

His administration has announced plans to encourage private companies to set limits on their own carbon emissions. It has issued a climate-change research strategy document and invited feedback on it from the scientific community. It says that the United States will rejoin the international ITER consortium to build a magnetic-confinement fusion experiment. It has backed a proposal to develop a 'hydrogen car', together with the infrastructure for this elusive vehicle. And last week it unveiled plans to build a coal-fired power station, the carbon emissions from which will be safely sequestered somewhere underground (see page 7).

What does this flurry of activity amount to? Looking at the few details that have been released, not a lot. The three big technical projects in this list could feature as imaginative components in a long-term energy strategy. But, ITER apart, they look decidedly half-baked. The power station, for example, will be built "with international and industrial partners" who have yet to be consulted, using technology that does not currently exist, to meet an objective that can play only a marginal role in alleviating climate change.

The principal objective here, it seems, is to generate a vague public impression that the president cares about energy and the environment, while discreetly assuring the president's true allies in the coal and nuclear industries that he is firmly on their side. An administration that prides itself on being honest and straightforward is, in this particular regard, proving instead to be anything but. ■





Radio days
Atacama telescope
array gets the
go-ahead
p4



Root cause
Could transgenic
maize lead to pests
resistant to Bt toxin?
p5



Virus scare
Human death
triggers hunt for
bird-flu vaccine
p6



One year on
Earth-observation
satellite celebrates
first birthday
p8

Astronomers urge NASA not to cut corners on Hubble successor

Tony Reichhardt, Washington

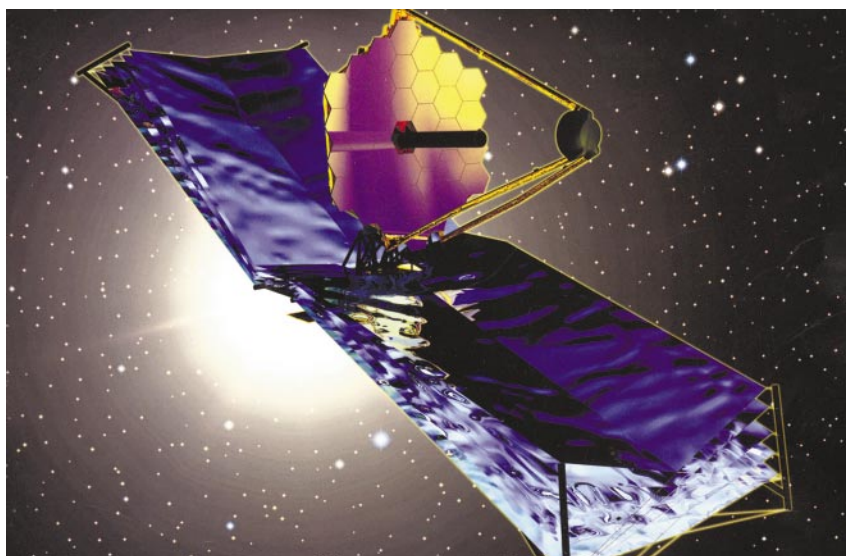
Researchers are rallying in defence of a key instrument that NASA is considering dropping from its next space telescope — the project ranked by US astronomers as their highest single priority.

Project scientists and congressional staff confirm that the space agency is considering axing the Mid-Infrared Instrument (MIRI)—one of three instruments to go on the James Webb Space Telescope (JWST) — in a last-ditch effort to keep the project on time and within budget. The JWST, formerly known as the Next Generation Space Telescope, is due to replace the spectacularly successful Hubble Space Telescope in 2011 as the premier observatory in space.

But ditching MIRI would be a “terrible scientific loss”, says Christopher McKee, an astrophysicist at the University of California, Berkeley, and co-chair of the National Academy of Sciences panel that gave the telescope top ranking in its 2000 decadal review. Others agree that MIRI should be preserved, even if it means delaying the telescope’s launch or cancelling other projects to pay for it.

The JWST has risen steadily in price from its original — and, some would say, unrealistic — target of \$500 million in the late 1990s to \$1.6 billion and going, not counting launch. Last November, NASA asked project managers, scientists and the telescope’s builder, Northrop Grumman Space Technology, to cut costs to stay within the budget. But no obvious solutions have surfaced apart from delaying the launch or reducing the telescope’s capability. “We still have a problem,” says Eric Smith, programme scientist for the JWST at NASA headquarters.

The James Webb telescope, named for the administrator who guided NASA during the Apollo Moon landings, will have a mirror 6 metres in diameter, over twice that of Hubble’s mirror. It will observe in the infrared waveband, where it can detect evidence of the oldest stars and galaxies in the Universe. A US-built instrument called NIRCам (Near-Infrared Camera) will cover infrared wavelengths from the edge of visible light up to 5 micrometres. MIRI — whose cost is shared by the United States and Europe — would



The James Webb Space Telescope could lose one of its three instruments unless more cash is found.

extend the range to 28 micrometres, and a third European instrument called NIRSpect (Near-Infrared Spectrograph) would complement NIRCам by taking spectra of stars and galaxies in the same infrared wavelengths.

Although NIRCам is the telescope’s prime instrument, it needs MIRI’s corroborating observations to answer some key scientific questions. At longer wavelengths, for example, it will have trouble distinguishing between the Universe’s very first stars and those born later. MIRI will also allow studies of objects in the distant Kuiper belt in our own Solar System and of planet-forming disks around other stars.

In a letter sent on 29 January to Anne Kinney, head of NASA’s Astronomy and Physics Division, McKee and panel co-chair Joseph Taylor of Princeton University called MIRI “an essential part of the complement of instruments that will enable JWST to have a revolutionary impact on astronomy and astrophysics”. They urged NASA to consult with the science community before making any decision, and if necessary, to consider “decoupling or delaying other projects in order to ensure the success of JWST”.

Ironically, just as rumours were circulat-

ing last week that NASA might cut its \$100-million investment in MIRI, the European Space Agency’s Science Programme Committee gave approval for its half of the instrument, and agreed that the agency should provide an Ariane launcher for the telescope. Smith calls the Ariane offer “fantastic news”, but the project would still be short of at least \$100 million. And although Smith says NASA has reached no decision, MIRI is the obvious place to cut if the telescope is to stay on time and on budget.

Last month a NASA advisory subcommittee suggested delaying the launch by up to two years to save MIRI. In a letter sent on 2 February to the head of NASA’s science advisory committee, subcommittee chairman Alan Dressler of the Carnegie Observatories wrote that “MIRI is critical for mission success”.

NASA is loathe to delay the JWST, as this would increase the cost of the project and give ammunition to critics of its project-management skills. Embarrassed by budget overruns on the International Space Station, the agency badly needs to demonstrate fiscal responsibility. Rather than ask for more money, says Smith, “we’re going to solve the problem within the space science office”. ■

'Revolutionary' telescope gets green light

David Adam, London

Astronomers have finally got the official go-ahead to build the world's most powerful radio telescope, in the Chilean Andes.

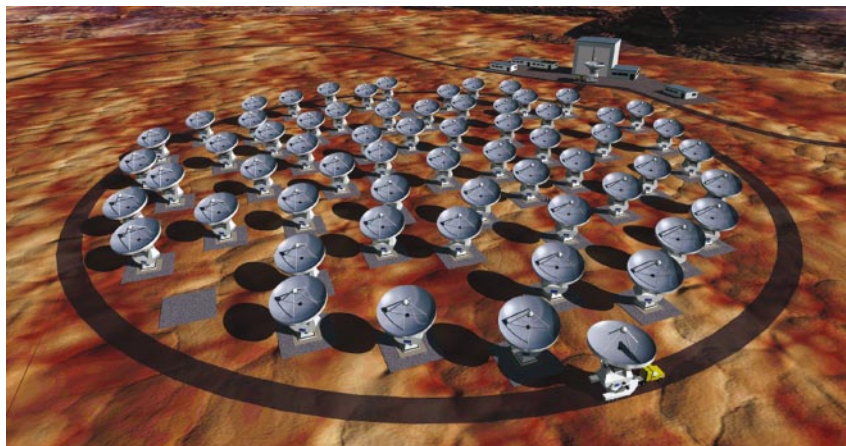
The \$650-million Atacama Large Millimeter Array (ALMA) is expected to help researchers explore processes such as the formation of galaxies, stars and planets, the intricacies of which are obscured from the most powerful optical telescopes by dust clouds.

But millimetre-range wavelength radiation will pass through these clouds and be detected by ALMA. The telescope will use 64 portable dishes spread over an area 14 km across to observe it with a spatial resolution up to 50 times higher than the most powerful current instruments, such as the IRAM array at Plateau de Bure in the French Alps.

"It's going to be revolutionary," predicts Peter Shaver, an astronomer with the European Southern Observatory (ESO), based in Garching, Germany. "Even our largest optical telescopes cannot see through the dust, so ALMA will allow direct observations of planet, star and galaxy formation for the first time."

ALMA will analyse higher radio frequencies than the existing Very Large Array, a radio telescope array in New Mexico, which observes cosmic events such as black holes. It will also pick up signals in a different part of the electromagnetic spectrum from the planned James Webb Space Telescope (JWST), which will observe infrared radiation (see page 3).

And unlike the JWST, ALMA's completion



A veritable array: ALMA's 64 antennas will allow the first direct observations of planet formation.

is fully assured: at a meeting in Washington DC on 25 February, representatives of the ESO and the US National Science Foundation signed the official agreement to build it. Both had already said that they would reach the agreement, "but it's still a relief to have it signed and sealed", says Richard Kurz, who is European project manager for ALMA and is also based at the ESO's Garching headquarters. Construction will now begin later this year, with the first scientific results due by the end of 2007.

One signature missing is that of Japan. Although Japanese astronomers had hoped to participate in the project from the outset (see *Nature* **401**, 627–628; 1999), their government has not yet committed any funding. If this situation changes, the current

'baseline' plans for ALMA could be upgraded, Kurz says. "Extra Japanese funding would allow us to add on some additional features," he explains.

ALMA will also help Chile to continue to develop local astronomy expertise: its scientists will receive 10% of the available viewing time on the telescope, and the project partners say that they will pay student fellowships to allow young Chilean researchers to work with the new facility.

Ultimately, the team hopes that the ALMA array will be able to monitor radiation with wavelengths of 0.33–10 mm, divided into 10 convenient bands. For now, instruments operating in just four of these bands are planned, but Kurz says that others can be added later. ■

Climate studies hold key to future of desalination plant

Rex Dalton, San Diego

Palaeontologists and hydrologists are being called in to help the US government decide what to do with a disused, \$400-million desalination plant on the banks of the Colorado River.

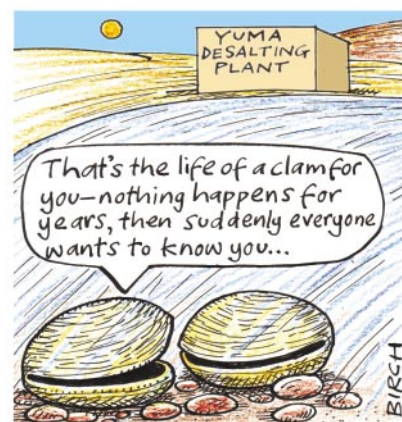
The researchers will study tree rings and clam shells, among other things, in a bid to establish long-term historical climate patterns in the river basin. These will help the US Bureau of Reclamation to determine whether it is worth reopening the Yuma Desalting Plant, which was designed to remove salt from the river before it reaches Mexico.

The plant was completed in Arizona in 1989, after salty water from the Colorado River had upset agricultural and natural ecosystems in Mexico. The salty water had entered the river from

irrigation run-off from agricultural fields, particularly in western Arizona. Other steps taken to address the problem included building a canal to divert agricultural run-off to a marsh near the Gulf of California.

The desalination plant was operated for a test period of about six months in 1992. But in January 1993, heavy rains diluted the river, making desalination operations unnecessary. The plant was subsequently criticized in media reports as a waste of taxpayers' money, and environmental groups complained about the potential impact of brine extracted by the plant that was sent to the Mexican marshland.

But a four-year drought in Arizona has led officials to try to find new ways of securing more usable water, while



meeting the requirements of the complex laws and treaties covering the river. A treaty signed in 1944 requires the United States

Agency 'ignoring its advisers' over *Bt* maize

Jonathan Knight, San Francisco

A strain of maize that is genetically modified to fight rootworm — a major crop pest — has won approval from the US Environmental Protection Agency. But scientists who were consulted before the 25 February decision say that the agency ignored their advice and is doing too little to ensure that insects don't develop resistance to the insecticide produced by the plant.

Last October, a scientific review board recommended that the strain should only be grown if farmers plant an equal area of non-transgenic maize next to it. Such a stipulation would have undermined the commercial viability of the strain, however, and the EPA has rejected it, saying that a 20% "refuge" of non-transgenic maize will suffice.

The decision has drawn immediate fire from members of the review board.

"They are putting the issue of sustainability on the back burner," argues Fred Gould, an entomologist at North Carolina State University in Raleigh.

The new *Bt* maize — developed by Monsanto of St Louis, Missouri, to express an insect toxin gene from the bacterium *Bacillus thuringiensis* — differs from existing varieties in several ways. The *Bt* maize that is already widely planted in the United States targets the European corn borer, a moth whose larvae eat stalks and leaves, but is useless against the corn rootworm, a beetle that costs farmers US\$1 billion a year in pesticide treatments.

The various strains of *Bt* bacteria collectively produce some 40 classes of insect toxin,



Rootworm: could low-dose *Bt* create resistance?

and the new variety of maize incorporates one that attacks only the corn rootworm and its close relatives. But it also produces much less toxin than existing *Bt* crops — increasing the prospect that insect populations will have time to develop resistance to the toxin. To mitigate this risk, refuges of non-*Bt* varieties are planted so that susceptible insects can multiply. These then mate with resistant insects and dilute out

the resistance genes in the next generation.

But this strategy only works if the transgenic plants produce enough toxin to ensure that only a very small number of insects (which might spawn resistant offspring) survive, say Gould and other critics of the EPA decision. If the crop produces too little *Bt* toxin, the survivors will include a large number of partially resistant insects who are likely to find each other and enrich the gene pool for resistance, they say.

Whereas established *Bt* maize varieties produce high doses of toxin, the new variety kills only about half the rootworm larvae, according to data provided by Monsanto to the EPA. This is enough to help farmers, Monsanto claims, because it is just as effective as pesticide treatments and has the advantages of saving time and doing less damage to the environment.

But with such a low mortality rate, resistance is certain to arise, the strain's critics say — the only question is when. Farmers can delay resistance by planting larger refuges. So Gould and ten other members of a scientific review board that looked at Monsanto's application urged the EPA to require a refuge size of at least 50% of the total area planted with corn.

In its ruling, however, the EPA sided with three dissenting review-board members, and sanctioned the 20% refuge size that Monsanto had requested. EPA spokesman David Deegan says that the agency's own calculations suggest that even if the new variety were planted all over the United States, it would take between 7 and 16 years for resistance to become a problem. Monsanto expects its anti-rootworm corn to cover no more than 5 million acres — 6% of US cornfields — by 2005.

"The registration is just for three years, and I don't think resistance is going to emerge in three years," argues Richard Hellmich, one of the board's dissenters and an entomologist with the Agricultural Research Service, the research arm of the US Department of Agriculture, in Ames, Iowa. Hellmich's group is one of several that will be monitoring rootworm in *Bt* fields for signs of resistance during this time. Refuge requirements can be changed later if necessary, he adds.

But by then it may be too late, warns Margaret Mellon, who directs the Food and Environment programme at the Union of Concerned Scientists, an advocacy group based in Cambridge, Massachusetts. It is hard to get farmers to plant less of something if they like it, she argues, and once a resistant population of insects arises, it can be very hard to eliminate.

"We need to protect *Bt* as a valuable resource and not use it up," she says.

to send 1.85 billion cubic metres of water to Mexico via the river, with the salinity level not exceeding set limits. The desalination plant could make a partial contribution to this target by removing salt from part of the river's water flow — it can desalinate 92.5 million cubic metres of water per year.

Reopening the plant would cost about \$25 million, and annual operating costs are also about \$25 million, so bureau officials only want to do this if it will be needed for a sustained period. In January, they convened a group of researchers at the University of Arizona in Tucson to develop a model for predicting droughts and higher-precipitation periods for the next 50 years.

University of Arizona palaeontologist Karl Flessa told the meeting that

studying tree-ring records in forests in the Colorado River basin could allow him to track droughts, but not short periods of high precipitation, for the past 500 years.

Flessa and his colleagues are also examining the ratio of two oxygen isotopes in clam shells in the river delta to estimate historical higher-precipitation periods. The isotope ratio reflects the degree of salt water and fresh water to which a clam was exposed, as one isotope is more abundant in fresh water.

If the scientific studies show that it would be cost-effective to reopen the desalination plant, bureau officials expect such a move to face legal challenges from environmental groups. Just keeping the facility on stand-by costs about \$3 million a year.

Vaccine sought as bird flu infects humans

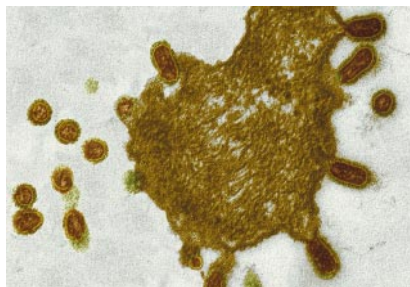
David Cyranoski, Tokyo

A bird flu virus is the subject of intense scrutiny amid fears that it could spark a human flu pandemic. Virologists increased their vigilance after the avian virus infected two people in Hong Kong last month.

Flu pandemics, which can kill millions of people, are thought to arise when viruses carried by animals — often birds — form hybrids with viruses that can transfer readily between humans.

It is difficult to predict when this will occur, virologists say, but the cases in Hong Kong — a 33-year-old man who died on 17 February and his 9-year-old son, who is in a stable condition — have focused attention on an avian flu strain called H5N1. Initial examination of the strain suggests that, at the moment, it cannot move easily from human to human. “Fortunately, this is purely avian,” says Malik Peiris, a microbiologist at the University of Hong Kong.

But health officials are nonetheless rushing to prepare a vaccine to control the spread of H5N1 among humans. Some are also



Virologists fear that the H5N1 virus, shown bursting from a cell, could trigger a pandemic.

alarmed about an inadequately understood outbreak of respiratory illness that has killed several people in the nearby Chinese province of Guangdong — although no link has been confirmed between this and the avian flu.

Flu viruses are constantly mutating, often exchanging DNA with other similar viral strains. It was hybrids between avian and human influenza viruses that led to the last two flu pandemics, in 1957 and 1968. H5N1 has a worrying track record — it was not

supposed to cross over from birds directly to humans, but in 1997 it infected a few hundred people in Hong Kong, killing six of them.

Moreover, the latest strain is genetically distinct from that of 1997. Between the two strains, only one gene — for the protein haemagglutinin — is similar. “Over the past five years, the H5N1 virus has been moving around southeast Asia picking up new genes,” says Klaus Stöhr, director of the global influenza network run by the World Health Organization (WHO) in Geneva.

Clearly, H5N1 can infect and kill people, so if it develops the capacity to move easily from human to human, it could conceivably cause a pandemic. The virus could gain this ability either by adapting itself through genetic mutations once inside the body, or by mixing with a widely circulating human flu strain, such as H3N2, in someone infected with both viruses, says Stöhr. “We have to watch it closely.”

The threat has pushed the WHO to accelerate its vaccine development. At an emergency meeting last week in Geneva, researchers discussed the problems involved in developing a vaccine against H5N1.

Flu vaccines are normally made by injecting viruses into hens’ eggs and letting them multiply as the embryo develops. But the H5N1 virus, being an avian flu, kills the chick embryos. So selected genes from the virus are instead inserted into a different flu virus before it is injected into the eggs.

If all goes well, a vaccine could be designed within four to six weeks if it is needed, says Ian Gust, director of the WHO collaboration centre for influenza at the University of Melbourne, Australia. Animal tests, quality control and manufacturing, however, would take several more months, raising the danger that a pandemic could occur before the vaccine was available to control it.

Many researchers suspect that the latest H5N1 strain originated in mainland China, which has led to increased interest in the respiratory illness sweeping Guangdong. China has so far not invited external parties to visit the province. “We would like to support China in terms of surveillance and disease definition, but we haven’t been invited in,” says Nikki Shindo, medical officer for the WHO’s global influenza programme.

Beijing has told the WHO that the epidemic was caused by the bacterium *Chlamydia pneumoniae* and is now under control. “But that can’t be the whole story,” says a Chinese researcher close to the investigation of the outbreak. “From a clinical standpoint, it seems to be related to a virus, and we cannot rule out the bird flu.”

Researchers say that the data do not exist to prove or disprove any link. “We don’t know,” says Gust. “We need careful molecular and epidemiological studies.”

India budgets for boost in research

K. S. Jayaraman, New Delhi

India has cheered its research community by announcing increases in research funding as part of a generally austere budget.

In the financial year that begins on 1 April, the government plans to spend Rs147 billion (US\$3.1 billion) on research and development (R&D) — an increase of about 9.5% — despite an economic slowdown that was exacerbated by the failure of the rains in last year’s monsoon.

“The drought may have axed some projects in other ministries, but not in ours,” Valangiman Ramamurthi, secretary of science and technology, told *Nature*.

“The jump is not as big as last year’s 25%,” says Ragunath Mashelkar, a chemical engineer who runs several of the ministry’s applied research labs, “but it is adequate”.

The rise reflects the government’s ambitious plan to expand India’s total R&D spending from 1.1% of gross national product to 2% by 2007, says Ramamurthi. The funding hike comes despite the fact that the government has allotted only Rs60 billion in new money to all of its ministries this year.

Ramamurthi points to injections of an extra Rs2.15 billion for multidisciplinary research projects and Rs1.5 billion for drug research as examples of the government’s priorities. The latter is being given priority, government officials say, so that India can

India’s R&D spending, millions of rupees

Item	2002-03	2003-04
Agriculture	14,490	15,110
Atomic energy	33,520	38,010
Information technology	5,020	5,020
Medical research	9,330	9,880
Ocean research	1,740	1,990
Dept Science and Technology	9,320	11,760
Industrial research	9,850	11,360
Biotechnology	2,210	2,730
Space	21,630	23,680
Defence	27,000	27,340
Total	134,110	146,880

Source: Government of India

develop its own drugs when the copying of foreign ones is outlawed by a new patent regime that is scheduled to come into force in 2005.

Biotechnology, information technology and renewable-energy research all rank highly in the budget, which also earmarks funds for a controversial plan to build a fast-breeder nuclear reactor near Chennai in Tamilnadu state.

“Overall, the budget is very good,” says Debi Sarkar, a biochemist at Delhi University. “I have much more money for my projects than when I started my career six years ago and I have no complaints.” But he adds that the main problem faced by Indian researchers “has never been with money, but with bureaucracy”.

Cancer risk prompts US to curb gene therapy

Erika Check, Washington

Children with severe combined immunodeficiency disease (SCID) should not normally be treated with gene therapy because of the associated cancer risk, a US advisory panel says. But it believes that there is insufficient evidence to justify halting similar trials for other conditions.

The treatment for SCID, a disease that disrupts the development of the immune system, had been the only success for gene therapy so far. But the Food and Drug Administration (FDA) suspended US SCID trials and then halted 24 similar trials in January, after two children treated for the condition in a French study developed leukaemia.

Following the ruling on 28 February by the FDA's Biological Response Modifiers Advisory Committee (BRMAC), most of these 24 trials are expected to resume, when each has undergone an FDA safety review.

The BRMAC's decision is largely consistent with that taken earlier this month by the Recombinant DNA Advisory Committee, which advises the National Institutes of Health on trials sponsored by the agency. But the new recommendation has a broader reach, as the FDA regulates all US clinical trials, including those that are privately sponsored.

Although the committee's decision that the therapy should not be used unless there is no other option was widely expected, it still marks a setback for the field of gene therapy.

"The impact of these cases has reverberated in the field in a lot of ways," says Daniel Salomon, outgoing chairman of the BRMAC and a gene-therapy researcher at the Scripps Research Institute in La Jolla, California.



Marina Cavazzana-Calvo has evidence that viral vectors (inset) used in gene therapy could cause cancer.

"The potential for a risk of cancer is going to influence everything we do."

The major question now haunting the field is why the children in the SCID trial — led by Alain Fischer at the Necker Hospital for Sick Children in Paris — developed cancer. Most of the available evidence points to the retroviral vector used to shuttle the corrective gene into the children's cells. This is significant because modified retroviruses are often used as the vector in gene therapy.

Scientists now need to find out whether using retroviral vectors in other diseases will also cause cancer — or whether the particular gene being delivered, or the disease being treated, somehow increased the risk for the French SCID trial.

At the 28 February meeting, Fischer's colleague Marina Cavazzana-Calvo showed preliminary evidence that the vector, rather

than the gene inserted into it, was responsible for the leukaemia. The evidence is far from definitive, but if it holds up, it would mean that any other retroviral vector could theoretically cause the same problem.

And although only two of ten children treated by Fischer have developed cancer, there are some signs that others could potentially develop the disease (see *Nature* **421**, 678; 2003).

In the meantime, private-sector backers of gene therapy are becoming skittish. More than one company has scaled back investment in retroviruses and other vectors that integrate themselves into human DNA.

"There was a lot of hype that there would be clinical therapies using gene therapy by now, and it's logical that there's been a retrenchment," says Salomon. "There are some major issues that the field has to address." ■



D. PATRICK/CORBIS SYGMA; J. C. REY/SPL (INSET)

Coal-fired power plant to bury issue of emissions

Hannah Hoag, Washington

The United States plans to build a coal-fired power plant that will sequester its carbon dioxide emissions deep underground.

The \$1-billion, ten-year project, called FutureGen, would produce electricity and hydrogen. Financed with federal, private and international funds, it will serve as a working prototype for clean-coal

technologies and power production, say officials at the US energy department.

No proposed location and few technical details were provided by energy secretary Spencer Abraham when he announced the plan in Washington on 27 February. A Carbon Sequestration Leadership Forum will be held in Virginia in June to assemble a consortium of industrial and international partners for the project, he said, and details will be planned subsequently. Officials say that the plant will initially sequester 90% of its carbon dioxide emissions, a figure that may rise to 100% at a later date.

The initiative is the latest in a string of recent announcements that seem designed to answer criticism that the Bush administration has failed to provide leadership in addressing climate change (see page 1).

Some scientists and environmental

groups have welcomed FutureGen as a forward-looking project that will help to address climate change in the long term. But critics charge that the scheme's benefits are decades away, and that it will divert money and attention away from short- and medium-term efforts to address the problem of greenhouse-gas emissions.

"The fact that it is not coupled with serious short-term efforts is an indication that this administration does not take climate change seriously," says Cutler Cleveland, director of Boston University's Center for Energy and Environmental Studies.

Uncertainties surround the feasibility of large-scale carbon sequestration. In principle, carbon dioxide could be contained in terrestrial or marine sites. Tests to achieve this are currently under way on a small scale at various sites around the world. ■



Burning issue: can coal tackle climate change?

K. R. SVENSSON/SPL

World water shortage predicted as pollution problem worsens

Paris Future water resources could be limited by pollution, says the team behind the most comprehensive review of the topic so far.

If pollution keeps pace with population growth, 18,000 cubic kilometres of fresh water — almost nine times the total amount used worldwide for irrigation — will be polluted by 2050. The report, which was due to be released on this week by the United Nations, says that the pollution is partly due to the 2 million tonnes of waste dumped into rivers and lakes each day.

But the review played down the likelihood that such shortages could spark conflict. The authors surveyed water-related interactions between countries over the past 50 years, and found that most resulted in water-sharing treaties or the building of dams. “Some of the most vociferous enemies around the world have negotiated water agreements,” says the report. The authors point to the example of agreements over the use of water from the Indus: tributaries of the river flow across the partly disputed India–Pakistan border, but the group that oversees control of the river has survived two wars between the countries.

♦ www.unesco.org/water

NASA decides that three's a crowd on space station

Washington Life on the International Space Station is set to become lonelier. NASA and its international partners decided last week to reduce the crew from three to two people in order to conserve food, water and other supplies while the space shuttle is grounded.

How much research a crew of two can do is unclear. In 2001, NASA was forced to reduce intended numbers on the station from six or seven to three in response to budget cuts. The agency's advisers said this would allow little significant science to be done (see *Nature* 418, 263; 2002). But NASA administrator Sean O'Keefe insists that the two astronauts will have time for research.

The next inhabitants, who will join the station in April or May aboard a Russian Soyuz capsule, are likely to continue some of the current crew's studies of physiological responses to weightlessness. Russian Progress spacecraft will be used to ferry supplies to the station while the shuttle is out of action.

Study breeds cautious optimism over vCJD

London Cases of variant Creutzfeldt–Jakob disease (vCJD) in Britain are on the wane, according to an epidemiological study.

Azra Ghani and colleagues at Imperial College, London, modelled the course of the

Islands in the pink

London Cool waters from the Atlantic Ocean swirl around the Balearic Islands of Mallorca and Menorca (far right) in this image snapped by Envisat, an Earth-observation satellite developed by the European Space Agency.

Envisat was launched on 1 March 2002 (see *Nature* 415, 825; 2002), and this picture is part of an exhibition of images from the satellite to celebrate its first year in orbit, currently on show at the British National Space Centre in Leicester. The thermal picture shows cooler waters in white and pink and the warmer waters of the Mediterranean Sea in blue. Understanding the mixing of such currents is an important part of ocean-circulation and climate research.



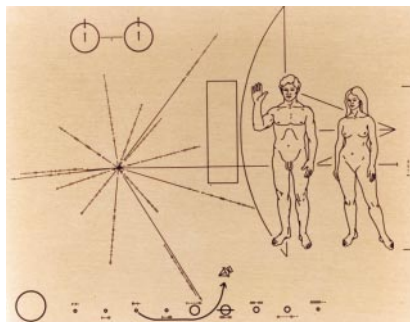
RUTHERFORD APPLETON LABS

neurological condition, thought to result from eating beef tainted with the protein that causes mad cow disease. They predict that the number of British people who may yet succumb is between 10 and 7,000 (A. C. Ghani *et al. Proc. R. Soc. Lond. B* doi:10.1098/rspb.2002.2313) — far fewer than the maximum number of fatalities predicted in 1998. The disease has killed fewer people each year for the past two years: 17 died in 2002, compared with 20 in 2001 and 28 in 2000.

But this is not necessarily a permanent decline. Another epidemiological study of vCJD, published last week in *The Lancet*, points out that all known victims have carried a particular version of the gene for the prion protein that goes awry in sufferers (N. J. Andrews *et al. Lancet* 361, 751–752; 2003). People with other genotypes may also be susceptible, but with longer incubation times.

Solar-System researchers in the dark as probe signs off

Washington In the end its plutonium batteries gave out, too feeble to send a radio signal to Earth. The Pioneer 10 spacecraft's final transmission, on 22 January, prompted nostalgic media coverage. But researchers studying the Solar System's outer reaches have lost what could have been a valuable data



Pioneer 10 won't be talking to us anymore, but the plaque it carries could be of interest to aliens.

source as the craft journeyed deeper into space.

Space scientists had hoped that Pioneer 10 would detect the termination shock — the region where charged particles from the Sun collide with magnetic fields and particles in interstellar space. The shock is thought to lie 80–90 astronomical units (AU) from the Sun (1 AU is the Earth's distance from the Sun). The spacecraft's twin, Pioneer 11, stopped transmitting in 1995. Just two spacecraft — Voyagers 1 and 2 — are now probing the region. Voyager 1 is now 87.5 AU from the Sun, slightly farther than Pioneer 10, whereas Voyager 2 is 69.5 AU away. None appears to have hit the boundary, although some data from Pioneer 10 are yet to be analysed.

The latest genomics, straight to your phone

Stockholm Biologists can now act on inspiration whenever it strikes, thanks to a new database that sends information about genome sequences to mobile phones.

“I was sitting on the bus reading a paper and I wanted to know what DNA sequences were available for the species I was looking at,” says the service's creator, genomics researcher Björn Ursing of the Karolinska Institute in Stockholm, Sweden. “Then I looked at my mobile phone.” Ursing went on to develop WiGID — the Wireless Genome Information Database. WiGID contains information on genome size, publications and scientific descriptions of the 117 organisms that have had their genomes sequenced.

Some researchers have questioned whether biologists need such information on their phones, but Ursing says he has already heard of scientists using the database during seminars. WiGID will complement BioWAP, a service developed at the universities of Turku and Tampere in Finland, which allows researchers to retrieve information from web-based genome and protein databases.

♦ wigid.cgb.ki.se

♦ www.uta.fi/limt/bioinfo/BioWAP

The coast road

America's first inhabitants were people from Asia who migrated over a now-submerged land bridge between the two continents. But when did they come, and where did they go after making their crossing? Rex Dalton reports.

Along the storm-lashed coast of southeast Alaska, where dense forests shroud weather-pocked limestone formations, Timothy Heaton spends his summers prowling remote islands in search of a special cave. He drops into sinkholes, squirms through narrow tunnels, and probes the sediments of underground caverns.

The object of his dirty, challenging quest is a cave containing specimens that will answer one of the great mysteries surrounding the peopling of the Americas: the date at which the first human colonists crossed from Asia over a land bridge that now lies submerged beneath the Bering and Chukchi Seas, and where the travellers went after that.

Heaton, a vertebrate palaeontologist at the University of South Dakota at Vermillion, is one of a growing number of experts who believe that the first Americans migrated along the Pacific Coast among the islands and bays of Alaska and Canada, at a time when the North American interior was an inhospitable, ice-covered wasteland. Recent discoveries have cast doubt on the conventional wisdom that North America was first colonized by a culture of big-game hunters called the Clovis, who ventured south across the continent's central plains only after the ice sheets retreated, reaching what is now the southwestern United States at least 11,500 years ago.

Now there is an explosion of interest in studying the climatic, environmental and geological conditions that prevailed along the Pacific Coast during the past 35,000 years or so. Investigators have high hopes of finding physical traces of early coastal migrants,

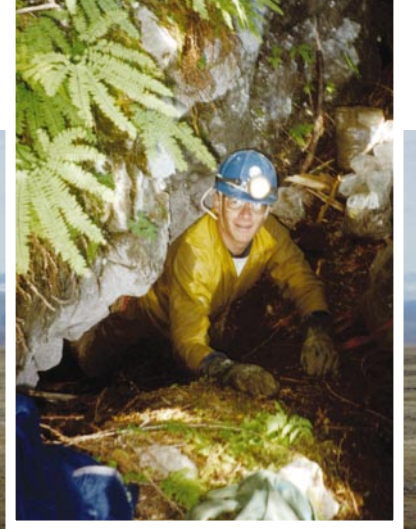
some of whom may have preceded the Clovis by many thousands of years. "We are just ramping up to nail down issues about the coastal route," says Julie Brigham-Grette, a geologist at the University of Massachusetts at Amherst.

Already, Heaton and his colleagues from the Institute of Arctic and Alpine Research (INSTAAR) at the University of Colorado, Boulder, have found a man-made tool that they have radiocarbon dated to about 10,300 years ago in a cave on Prince of Wales Island, southeast Alaska¹. They have also found bear specimens dating back some 41,000 years² — which is tantalizing, because where bears can live, so can people.

Who came first?

Archaeological discoveries made in New Mexico in the 1920s — in particular, distinctive fluted projectile points dating from 11,500 years ago — identified the Clovis as the first known Americans³. The timing seemed to make sense: before the ice melted, no one would have been able to migrate across the continent's frozen interior. And until about 2,500 years before the Clovis arrived in New Mexico, ice sheets extended throughout Canada and into the northern United States.

But more recent findings in South America have demolished the status of the Clovis as the original American pioneers. A quarter of a century ago, anthropologist Thomas Dillehay, now at the University of Kentucky in Lexington, made an astonishing claim: that a group of hunter-gatherers was well established at a site called Monte Verde in Chile 12,500 years ago⁴. For many years, the radio-



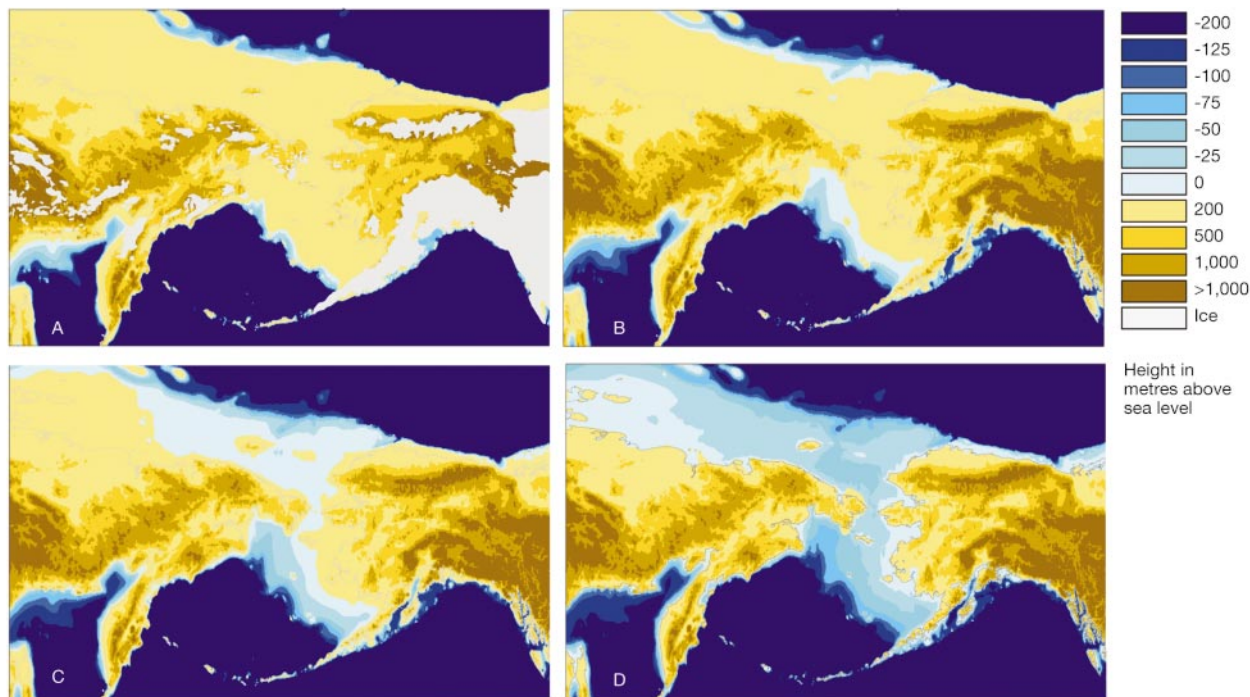
Going underground: Timothy Heaton hopes that his summer caving expeditions in Alaska will shed light on the route taken by the first Americans, who hiked across a desolate strip of tundra thousands of years ago.

carbon dating of Dillehay's specimens was disputed. But in 1997, Dillehay published a detailed monograph on the Monte Verde site⁵ that silenced his critics once and for all.

If people had migrated almost to the tip of South America by 12,500 years ago, experts agree they must have begun trekking south from Alaska before the glaciers retreated from the American interior. So in the past few years, attention has turned to the coastal route, where the moderating influence of the Pacific Ocean may have kept conditions fit for human habitation. "We are seeing a major paradigm shift," says James Dixon, an archaeologist at INSTAAR who has investigated caves with Heaton.

Confirming the coastal route is no easy matter, however. The terrain is forbidding, harsh winters restrict field studies to the summer months, and much of what was then the Pacific Coast of North America is now underwater. When the glaciers finally melted, the sea level rose by some 125 metres, covering both the land bridge from which the first American colonists migrated from Asia — a strip of tundra some 1,000 kilometres wide known as Beringia — and regions that now lie on the coastal shelf off southeast Alaska and British Columbia.

Last summer, a team of 20 scientists from five US institutions began the most ambitious study yet of Beringia. Led by Brigham-Grette, together with Lloyd Keigwin, a palaeoclimatologist at the Woods Hole Oceanographic Institution in Massachusetts, and Neal Driscoll, a geologist at the Scripps Institution of Oceanography in La Jolla, California, they sailed aboard the *Healy*, a new icebreaker operated by the US Coast Guard, to the Bering



Bridging the gap: calculations of sea levels 18,000 years ago (top left) and 10,200 years ago (top right) show that a land bridge existed between modern Siberia and Alaska. By 8,900 years ago (above left) this land was submerged, creating today's Bering Strait (above right). Quoted ages are calibrated for consistency with archaeological radiocarbon dates, as given in the text.

and Chukchi Seas (see map, below right). There, the researchers collected cylindrical sediment samples, or cores, from the now-submerged land bridge and the neighbouring ocean floor. Previously, the Chukchi Sea's persistent ice had prevented studies. But in the event, unusually warm conditions meant that the *Healy's* ice-breaking capabilities were not required. "It was embarrassing," confides Brigham-Grette. "We waited for an ice cutter and there was no ice."

Steaming back and forth across the region over two cruises, each lasting three weeks, the *Healy's* scientific crew pulled more than 100 cores from the sea floor. Some were at depths of 40–120 metres below the surface, on the former land bridge, whereas others came from depths of up to 2,500 metres in the deeper waters to the north and south. Now stored at Woods Hole, the cores will be subjected to extensive analysis to ascertain the climate, environmental conditions and sea level over the past 20,000 years. Research on the samples should determine, for instance, what portions of the land bridge were above water at what times.

Selecting sites for drilling, and obtaining intact cores, required some sophisticated technology. The scientists were particularly interested in sampling from former river valleys that are now submerged on the coastal shelf in the Chukchi Sea. But the valleys are overlaid by marine sediment, and are therefore hard to spot. Extracting intact samples

from both marine and river sediments is also challenging because of their high sand content — which means that the cores are likely to fall apart when pulled from the ocean floor.

A state-of-the-art sonar device, called SUBSCAN, allowed the team to find the submerged river valleys. SUBSCAN — a bat-winged affair the size of a desktop — was developed by Driscoll, together with Steven Schock of the Florida Atlantic University in Boca Raton and a company called EdgeTech in the same city. Towed behind the *Healy* some 40 metres above the sea floor, SUBSCAN emits an acoustic pulse that penetrates the sea floor to a depth of 40 metres in sand, and to 100 metres in clay or silt. The device features an advanced system for analysing the reflected sound that allows it to build up an image of the sedimentary layers, from which the researchers can discern the difference between river and marine sediments. "The resolution of the images produced is just dynamite," says Brigham-Grette, who notes that they clearly reveal the elusive V-shaped river valleys.

To ensure that sandy cores didn't disintegrate, the *Healy* researchers used a device that was developed at the Coastal Carolina University in Conway, South Carolina. The apparatus is named 'Taz' after the Warner Brothers cartoon character, because of its dervish-like whirling action. After being lowered to the sea floor by a cable, Taz's two

counter-rotating electric hammers produce a vibrating motion that forces a tube down into the sediment without disturbing the captured material. In conventional methods, a heavily weighted tube is driven into the ocean floor and pulled up with the core inside. But a sandy core has a tendency to fall out of the tube, so the Taz tubes feature 'fingers' at the bottom and also use suction to hold the sediment intact inside.

Although some 20 cores were pulled from



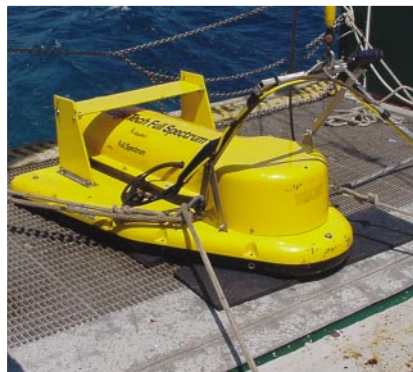
The SUBSCAN sonar device (right) pinpointed V-shaped river valleys in the Chukchi Sea floor.

Beringia's river valleys, there was a cost. 'Big Taz', the larger of the two devices, was lost after its cable snapped when the *Healy* was buffeted by a rough swell. It was a bitter blow for the Coastal Carolina team: "Those guys were almost crying," recalls Driscoll. But the researchers rallied, and welded a spare vibrating power head to a core tube to create a makeshift rig.

The cores will now be dissected to piece together a picture of life in Beringia. Pollen, other plant remains and beetles will be identified, and the specimens will be radiocarbon-dated. Various collaborators with special expertise will be brought in to assist the effort. For instance, Scott Elias, a palaeontologist at Royal Holloway, a college of the University of London in Egham, west of the city, will be looking for samples of beetles from lake sediment or peat bogs, which should reveal clues about Beringia's past climate and habitats.

Data from the *Healy* cruises should reveal when people would have been able to cross the land bridge, providing valuable background information for Heaton and his colleagues as they continue their search for signs of human activity in southeast Alaska. Heaton, a keen amateur caver, began exploring the region's caverns in 1991, and two years later struck lucky. In the north of Prince of Wales Island, he ventured into the 'On Your Knees' Cave — so named because it can only be entered by crawling through a narrow tunnel — which turned out to contain a host of specimens. Subsequent years of summer digs by Heaton and his archaeologist colleagues have unearthed the oldest signs of human habitation in the Pacific Coast region, from sediments in the cave's floor.

The specimens include a bone tool that has been radiocarbon dated to 10,300 years ago¹, a human bone dated at 9,200 years old⁶, and blades made from obsidian — a volcanic glass found in lava beds — of the same vintage⁷. The latter have shown the ancient inhabitants of Prince William Island to be



relatively well-travelled coastal seafarers. Craig Lee, an anthropology doctoral student now at INSTAAR, has used X-ray fluorescence spectroscopy to examine trace elements in the obsidian, and so determine its source. His studies show that the material in On Your Knees Cave came from Suemez Island, about 150 km to the south, and Mount Edziza, nearly 400 km to the north-west in mainland British Columbia⁷.

First footers

In themselves, these findings don't explain how people got to Monte Verde so early, because they postdate Dillehay's Chilean discoveries. But Heaton is confident that much earlier signs of human settlement along the Pacific Coast of North America are just waiting to be found. "This was not the first guy in the area," he says of the single human bone uncovered so far. Indeed, his confidence received a major boost in 1994 with the discovery of the 41,000-year-old bear bones² in On Your Knees Cave.

Still, a significant problem remains, says Heaton: "So many of the areas where people roamed are underwater now." The key may



Tell-tale tool: this bone implement proves that humans lived in Alaska at least 10,000 years ago.

be finding more caves like On Your Knees, which was exposed on the coast at the time of the initial southward migration and remains above sea level today. In this search, the work of Renee Hetherington of the University of Victoria in British Columbia could be crucial.

Hetherington studies molluscs, and last August in Anchorage, Alaska, at the seventeenth biennial meeting of the American Quaternary Association, she told how mollusc specimens, along with geological records, can point to areas where people could have survived during the most recent glaciation. As part of her doctoral studies, completed last spring, Hetherington investigated the region around Queen Charlotte Sound off the British Columbia coast — just south of where Heaton has been exploring.

Species of mollusc found in coastal waters — including clams, oysters and mussels — have certain environmental preferences, such as water depth, turbidity, salinity, temperature and sediment conditions. By carbon-dating mollusc shells, and studying the species present and the sediments in which they were found, Hetherington was able to determine the sea level and the positions of coastlines some 14,250 years into the past. Comparing her reconstructed map for this period with today's coastline revealed that hundreds of kilometres of the current British Columbia shoreline may have been exposed at the time when the migrants who eventually made it down to Monte Verde were passing through. "The trick is to determine an environment that is suitable for humans," says Hetherington. "The mollusc record can do that."

This June, Heaton will be returning to a promising cave he found in 2001 on Coronation Island, west of Prince of Wales Island. But he also plans to check Hetherington's maps for new targets to explore.

For Charles Schweger, an anthropologist at the University of Alberta at Edmonton, Canada, the convergence of the work of Heaton, Hetherington and the *Healy* crew is a welcome development. "People have languished far too long on the interior route," he says. "There are great new data and new ways to look at them."

Definitive evidence for the idea that America's first colonists migrated along the Pacific Coast during the most recent Ice Age may not yet have emerged, but Heaton suspects that there won't be a long wait. "It's just a matter of time," he says. ■

Rex Dalton is *Nature's* US West Coast correspondent.

1. Dixon, E. J. *Quat. Sci. Rev.* **20**, 277–299 (2001).
2. Heaton, T. H. *Curr. Res. Pleist.* **12**, 92–95 (1995).
3. Haynes, C. V. *Science* **145**, 14098–14113 (1964).
4. Dillehay, T. D. *Sci. Am.* **251**, 106–117 (1984).
5. Dillehay, T. D. *Monte Verde: A Late Pleistocene Settlement in Chile Vol. 2* (Smithsonian Inst., Washington DC, 1997).
6. Dixon, E. J. *et al. Geoarchaeology* **12**, 689–712 (1997).
7. Lee, C. M. in *32nd International Arctic Workshop of the Institute of Arctic and Alpine Research* 120–121 (INSTAAR, Boulder, Colorado, 2002).



The lab arm of the law

Forensic science is already a mainstay of modern police work. But are criminologists missing a trick by failing to apply the latest scientific findings to crime prevention? Jim Giles investigates.

I could walk into any laboratory and generate a crime-prevention application from their last few papers,” claims Ken Pease, a criminologist who is a visiting fellow at University College London (UCL).

For many in Pease's discipline, the idea that biologists, chemists or physicists have much to offer the field is unfamiliar. But a small band of criminologists is now arguing that science can be central to preventing crime. “It's about giving scientists a dirty mind,” says Paul Ekblom, a criminologist at the Home Office, the British government department responsible for criminal justice and law enforcement. “Getting them to think ‘thief,’” he says — or ‘rapist’, or ‘terrorist’, for that matter.

The Jill Dando Institute of Crime Science, based at UCL, exemplifies this fresh approach. Dando, a British television presenter, was gunned down at her London home in 1999 in an apparently motiveless attack for which Barry George, a local man with a history of mental problems, was later jailed. Colleagues and friends of Dando — who presented a show called *Crimewatch*, which appeals for public help in solving crimes — raised funds for an institute to be formed in her memory. It opened its doors in spring 2001.

At the helm is Gloria Laycock, who previously worked as a psychologist in prisons and for the Home Office. “The institute is a reaction against traditional criminology,” she says. “Criminologists look for social causes of crime. A crime scientist would ask what science has got to offer.”

Crime scientists recast criminals as ordinary people reacting to a situation in which a crime can be committed, rather than as individuals driven to deviance by their social circumstances. About a third of all British men in their mid-forties have a conviction for at least one offence, Laycock points out. Against this background, scientific methods can be used to analyse the cir-

Gloria Laycock is putting science in the spotlight in the fight against crime.



COREIS SYGMA

About a third of all British men in their mid-forties have a conviction for at least one offence.

cumstances under which crimes are committed, and to seek practical means to change those circumstances. For those who find criminology too ideological, crime science sounds like a *Realpolitik* alternative.

This approach has not endeared Laycock to some criminologists, who accuse her of neglecting crime's social origins. But her institute's researchers — 19 have been recruited so far, including visiting fellows — are bubbling over with speculative ideas. Pease, for example, wonders whether we might learn about the best way to deploy closed-circuit television (CCTV) cameras by studying the strategies used by groups of foraging primates to watch out for predators. Laycock speculates that mathematicians working on network

theory could help to tease patterns from crime statistics that might prove useful in devising prevention strategies.

Some scientific disciplines — computer science, for instance — already have a track record of contributing to crime prevention. Sergio Velastin of Kingston University in Surrey, near London, is currently testing a software system that could help CCTV operators spot problems on the London Underground, from packages that might contain bombs to people who are considering committing suicide. The software aims to take the pressure off CCTV operators, who often miss such things because they have to monitor large numbers of screens simultaneously. “We know they only look at 10% of cameras at any given time,” says Velastin.

The software starts by focusing on one area for up to an hour. This helps it to distinguish fixed areas of the image that it can ignore, such as the floors and walls. It can then spot suspect packages in the area of the image more easily. “There is always lots of movement,” says Velastin. “The biggest clue is something that is not the background but is not moving.”

The system can also be primed to spot

potential suicides. Train drivers and CCTV operators have learnt to recognize the movements of people who are thinking about jumping in front of a train. Such people, says Velastin, tend to position themselves at the end of the platform, stay there while several trains pass by, and often walk up to and away from the platform edge. Tested at a London Underground station last August, the system spotted 90% of events of interest and raised false alarms 2% of the time. Importantly, CCTV operators said that they were happy working with it.

Family conflict

Laycock, Pease and their colleagues are even more interested in drawing in researchers from disciplines that haven't traditionally been seen as relevant to the fight against crime. So far, they have made only a few firm links with UCL's mainstream science departments — recruitment and fundraising have taken up most of the Jill Dando Institute's first year. But as an example of the type of work that crime scientists hope to promote, Ekblom points to the research of evolutionary psychologists Martin Daly and Margo Wilson of McMaster University in Hamilton, Ontario, who in the 1980s began to challenge traditional ideas about homicides within families.

Sociologists have often asserted that violence is far more common between family members than between those who are not related. But Daly and Wilson thought that this idea was strange — evolutionary theory says that animals sharing genes by common descent should be less inclined to enter into conflict with relatives, not more so.

Controlling for the many different factors involved in homicide is difficult. For example, family members are more likely to fight each other simply because they have more opportunity to do so. So Daly and Wilson analysed a homicide data set collected by Detroit police, which provided information on whether the murderer and victim lived together. Taking just the latter cases, the researchers found that genetically unrelated people who shared accommodation were eleven times more likely to kill one

another than co-residing blood kin. This higher homicide rate applied equally to married couples and individuals without any relationship other than the fact that they shared accommodation¹. Significantly, Daly and Wilson also showed that stepchildren were at particular risk of being killed by their step-parents² —

Ken Pease seeks inspiration from scientific papers.



Alternative TV: the appliance of computer science can enhance CCTV's ability to identify problems.

a finding that might help social workers prioritize their registers of 'at risk' children.

Such work provides some support for the idea that there is untapped crime-prevention potential within mainstream science. But is Pease's confidence really justified? *Nature* decided to call his bluff, and presented him with six recent papers from UCL labs, selected by arbitrarily picking two departments and noting the three most recent papers listed on their web pages. We gave him two weeks to come up with crime-prevention ideas from at least two of the papers. Pease happily accepted — although he was a little flummoxed when the manuscripts arrived in his inbox. "My first reaction was: 'Oh shit!'," he admits.

Arresting ideas

One paper³, from climate researchers Mark Saunders and Budong Qian, examined how records of sea surface temperatures in the north Atlantic could be used to predict the North Atlantic Oscillation — the pattern of high and low sea-level air pressure that dominates winter weather around the region. Pease admits to not having a specific application for this paper, but notes that some studies have suggested that burglary rates fall during cold weather. The paper by Saunders and Qian made him wonder whether more effort should be put into investigating the links between crime and variations in seasonal weather patterns, in the hope that this might help in deploying police resources.

Another paper⁴, from Sham Kakade and Peter Dayan of UCL's computational neuroscience unit, compared theoretical models of animal learning. The models make use of variables such as the rate at which a particular reward is obtained, in this case the number of times the animal received food over a set time interval. Pease was struck by the fact that similar variables are used to compare

crimes. Burglary, for example, has a lower rate of reward than shop theft, as the latter is easier and quicker to carry out. But the variability of the reward from shop theft is low as it is unlikely to yield a big return.

Studies of criminals in New York have shown that groups of illegal squeegee merchants — people who demand to be paid for washing car windows at traffic lights — contain a high proportion of armed robbers⁵. Pease speculates that criminals sometimes choose to combine crimes with a high rate of reward and low variability — such as illegal car washing or shop theft — to one with a low rate of reward and high variability, such as armed robbery or burglary. If so, asks Pease, might these choices be modelled using the learning theories studied by Kakade and Dayan? A better understanding of the links between different kinds of crime might help law-enforcement agencies to devise crime-prevention strategies, he argues.

Dayan, however, declined to comment on Pease's speculations about his work. Does this indicate that the Jill Dando Institute will face some scepticism from UCL's mainstream science departments as it tries to interest them in crime-prevention applications? As Laycock, Pease and their colleagues begin to ramp up their efforts to bridge the cultural divide between science and criminology, this question will be under the magnifying glass.

Jim Giles is *Nature's* associate news & features editor.

1. Daly, M. & Wilson, M. I. *Am. Anthropol.* **84**, 372–378 (1982).
2. Daly, M. & Wilson, M. I. *The Truth About Cinderella: a Darwinian View of Parental Love* (Weidenfeld & Nicolson, London, 1998).
3. Saunders, M. A. & Qian, B. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL014952 (2002).
4. Kakade, S. & Dayan, P. *Psychol. Rev.* **109**, 533–544 (2002).
5. Bratton, W. *Turnaround* (Random House, New York, 1998).

Jill Dando Institute of Crime Science

♦ www.jdi.ucl.ac.uk

The subtle beauty of art in the service of science

An illustration may be intended to emphasize details, convey an idea or raise questions.

Sir — As a professional scientific illustrator I feel compelled to respond to Julio Ottino's Commentary "Is a picture worth 1,000 words?" (*Nature* **421**, 474–476; 2003). I believe that Ottino's criticisms of scientific illustration are founded on an incorrect understanding of the field.

Galileo's drawings can't be compared with magazine covers: they are two unrelated types of illustration. It is incorrect to conclude from such a comparison that scientific disparity exists between them because the magazine covers "are left in the hands of artists and illustrators" — this artwork was intended to enhance editorial material rather than to illustrate research.

Such conceptual illustrations are designed to pose questions. Their use on the cover of a science magazine offers the promise of articles that inform these questions. The cover art of the 30 January 2003 issue of *Nature* (see figure) and the related News and Views and Letter (*Nature* **421**, 489–490 & 530–533; 2003) follow this convention. The image does not illustrate the research itself; that is not its intended application. Furthermore, the choice of digital medium, whether used by the hand of a scientist or the hand of an artist, has no bearing on this question.

Scientific illustration follows a different mandate, and it can often be found within the pages of the very magazines under



discussion. These drawings outline structure and clarify detail, as required by the subject and requested by the researcher. Because they communicate subtleties and eliminate the ambiguities of language, scientific illustrations are an important, often necessary, element in precise communication (see *The Guild Handbook of Scientific Illustration*, edited by E. R. S. Hodges; Van Nostrand Reinhold, 1989).

Scientific illustration is a clearly defined field that benefits from active collaboration between scientist and illustrator. Using

their professional observational skills, scientific illustrators strive to render the most accurate representation of their subject. It is, by definition, art in the service of science. The act of drawing is, in essence, the act of editing. Complaints about omitted details miss this important point. Scientific illustrators are trained to eliminate non-essential information. The twisted stem of a dried plant is smoothed out. The broken edge of a fossil bone is repaired. Cracks and discoloration may be removed. These subjects are thus rendered in a way chosen to amplify those details that require emphasis.

Scientific illustrations, even conceptual cover art, should be as accurate as possible. However, Ottino's proposal to establish rules governing the use of realistic rendering techniques is superfluous. Professional standards are already in place for scientific illustration. Magazine editors recognize that their educated readership can distinguish between a beautifully rendered concept and the current state of scientific research. Scientific illustrations exist within this context. They communicate with and within conventions that reach back in time from this issue of *Nature* to the pages of Galileo's notebooks.

Frank Ippolito

Division of Vertebrate Paleontology, American Museum of Natural History, 79th Street & Central Park West, New York, New York 10024, USA

No strings attached to \$225m sponsorship

Sir — David Ritson in his Commentary "Fuel for thought" (*Nature* **421**, 575–576; 2003), addressing Stanford's new Global Climate and Energy Project (G-CEP), inaccurately describes the motivations and arrangements for this programme.

The premise of G-CEP is that energy is a critical component of modern societies. Supplying energy for a growing world population while significantly reducing emissions of greenhouse gases is one of the grand challenges of this century. It is entirely appropriate for Stanford faculty and students, working with institutions around the world, to engage in research that will have global benefits.

Stanford's independence is fully protected. The project's sponsors have their own substantial, related research programmes. They have chosen also to support research in a university because it brings a healthy independence of views that they value and

support. Academic freedom is an essential component of that, as both Stanford and G-CEP's sponsors agree. Stanford's considerable experience in working with companies, both in protecting the interests of research and in developing applications, will guide this project.

Any technologies developed under the G-CEP will be widely available, and all the work done will be reported publicly, as required by Stanford's standard policy on openness in research. Patents that arise from the research will be held by Stanford. During the first five years, a short period on the timescale of changes in energy systems, the sponsoring companies can license those technologies. Subsequently, Stanford and sponsors may license and sublicense broadly.

The G-CEP agreement explicitly calls for Stanford to explore broadly technology alternatives that can supply energy with substantial reductions in greenhouse emissions. There has been no attempt to favour fossil fuels or to exclude other energy sources. An initial G-CEP project

involves research on the biological generation of hydrogen, for example.

The three-year funding commitments under G-CEP are less restrictive than many research support agreements. Federally funded research projects often make continued support contingent on availability of funds and acceptable progress; funding for existing projects can be delayed or cancelled if agency budgets are reduced.

Under G-CEP, the proposal cycle will be shorter and the reporting requirements less stringent, and there will be greater flexibility to pursue new ideas.

Charles Kruger*, Franklin M. Orr Jr†

*Vice Provost and Dean of Research, Building 18, Stanford University, Stanford, California 94305, USA

†Director, Global Climate and Energy Project, Stanford University, Stanford, California 94305, USA

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

A champion of new technologies

Despite its strengths, India needs to invest far more to retain its lead.

Ashok Parthasarathi

Over the past 30 years, India has built up technological independence in many areas. The government's strategy of maintaining the science base has resulted in a steady increase in research and development (R&D) expenditure as a percentage of gross national product (GNP). Here I review India's many technological strengths, focusing on telecommunications, solar energy and supercomputers as examples. Although strong in these areas, India must undertake 'mission mode' R&D — technology development and commercialization programmes — to exploit its advantages and/or to respond to pressing national needs. It is particularly important for these initiatives to be undertaken as government–industry partnerships.

As long ago as 1938, India gave pride of place to science and technology. After independence in 1947, Prime Minister Nehru was fond of saying: "The future belongs to science and those who make friends with science." Ever since then, the promotion of science and technology and their use in social, economic, commercial and security spheres has continued unabated, with financial investment in the two steadily increasing.

Domestically developed technologies must be internationally competitive. But India has a barrier: the legendary reluctance of large-scale Indian industry — both public and private — to invest in and undertake R&D and innovation on a substantial and sustained basis. Government initiatives in the development of human resources, fiscal and financial policies, the building of institutional capacities and direct funding of R&D institutions and industry, provide hope that such objectives will be increasingly achieved over the next decade. But the journey will not be easy.

Industry's reluctance to invest adequately in R&D is a problem facing many 'intermediate' countries, including China. But developments in the 1990s seemed to suggest that India was pulling ahead. For example, in pharmaceuticals and agrichemicals, R&D-to-sales ratios of many of India's leading companies grew rapidly from less than 2% in 1990 to close to 6% last year. The same upward trend is seen in the number of international patents obtained by such companies. Solar energy is a particularly good example of an area in which there is both strong research and a high number of commercial products in use.

This improvement is a result of an investment initiative begun by the government four years ago. It selected strategic areas of indigenous technologies for development, includ-



Success symbol: the International Technology Park Building in Bangalore holds nearly 50 high-tech firms.

ing biotechnology, vaccines against prevalent communicable diseases, microelectronics, advanced batteries and technologies in atomic energy, space and defence.

The selection criteria varied. For example, biotechnology was chosen partly because of its impact on agriculture, health, industry and the environment, and partly because of heavy product patenting by foreign companies unwilling to license technologies to Indian companies. Vaccines, on the other hand, were picked because little R&D was being done in the industrialized countries for the tropical-disease market. Advanced batteries won their place for their potential to reduce costs and improve solar photovoltaic-based power sources (solar energy), as well as potentially enabling electric vehicles to become economically viable. Microelectronics, atomic energy, space and defence technologies won out because they were restricted by export controls.

Over the same four years, the government provided incentives for industry to invest in R&D. These included tax concessions; excise duty waivers on products covered by patents taken out in North America, the European Union or Japan; and an incentive scheme for 'R&D companies' undertaking in-house R&D and licensing the resultant technology.

Solar energy

In response to the price hike of crude oil in August 1973, and as a result of the vision of then Prime Minister Indira Gandhi, India began a solar-energy programme in 1975. This effort was reinforced when the Department of Non-Conventional Energy Sources was established in 1982, the first, and so far

only, government agency in the world for renewable energy.

India now has 9 manufacturers of solar photovoltaic cells and 20 manufacturers of photovoltaic panels, with an installed production capacity of 20 MW per year. This is the fourth largest capacity in the world and is by far the largest in a developing country. About 900,000 commercial photovoltaic systems have been designed, developed, manufactured and installed across the country to provide electrical energy that is desperately needed in rural, remote and industrial areas.

The energy provided is used for a wide range of essential services, including water pumps for drinking and irrigation, street lighting, refrigerators for storing vaccines in rural health centres, warning systems at unmanned rail-road crossings (of which India has more than 9,000), communication equipment, water desalination plants in desert settlements, anticorrosion protection systems for cross-country oil and gas pipelines, and the communication and control systems on offshore oil and gas platforms.

Telecommunications

In 1984, the government found that the telecommunication switching systems designed and engineered abroad were not really optimal for Indian applications and caused operational problems. So it set up the Centre for the Development of Telematics (C-DOT), an autonomous design and engineering organization aimed at developing digital switching systems.

This decision to develop a fully digital exchange came at a time when many international companies were still trying to sell

obsolete analogue systems to India. Over the next decade, C-DOT worked closely with the electronics industry to develop a family of switches, ranging from small 128-line systems designed to work in harsh rural environments, to international state-of-the-art, 40,000-line capacity exchanges for cities, all mutually compatible.

Distinctive features of the technology include the use of the most powerful software available, a wide base of component suppliers to feed the assembly lines, and a 'model' plant built just outside Bangalore. This plant was set up by the public-sector company Indian Telecommunication Industries to produce and test pilot exchanges in the telecom network under live operational conditions. High-quality technical documentation was generated to cover the engineering, manufacturing, testing, evaluation, installation and commissioning of the network, as well as the training of operating and maintenance engineers.

By the end of March 2002, about half of the 37 million lines in the Indian telecom network had C-DOT switches. C-DOT licensees have secured significant orders and exported switches to about 12 developing countries against stiff competition from multinational companies. This has been possible because C-DOT switches are not only technologically state-of-the-art, but are also 30% cheaper than those of its competitors.

Supercomputers

In 1989, recognizing the need for high-performance information technology, the government set up the Centre for the Development of Advanced Computing (C-DAC). This was a national initiative to design and produce supercomputer systems based on parallel-processing technology. C-DAC has since brought out three generations of PARAM supercomputers, which have increasingly advanced technologies and computing power, with an equivalent of US\$25 million in investment.

Typical scientific and engineering applications for C-DAC's supercomputers are in weather forecasting, seismic data processing, computational fluid dynamics, structural mechanics and bioinformatics. A particularly successful use, which is of crucial economic and social importance, is the modelling of the monsoon by the National Centre for Medium Range Weather Forecasting in New Delhi.

The price:performance ratio of C-DAC's supercomputers is better than those of any North American, European or Japanese supercomputers, in part because of the initial adoption of parallel-processing technology. The PARAM 10000 commissioned by C-DAC in 1998 was at that time the most powerful supercomputer in Asia outside Japan, with a computing power of 100 gigaflops. C-DAC is currently building an even more powerful system, PARAM 20000,



Dynamic solution: solar energy has brought much-needed power to remote parts of India.

with teraflop capability, planned to be in action by mid-2003. This will be used in climate and molecular modelling, genome sequencing, and two- and three-dimensional seismic data processing. C-DAC has supplied more than 50 supercomputers to domestic and export customers, including Russia, Canada, Germany and Singapore.

The next development is an 'iGrid' to link up the main supercomputing sites, providing researchers with access to a high-performance computing facility. This will also build up skills in the front-line area of supercomputing, which will have spin-offs in applications such as networking, security, visualization and large-database management. Locally developed and commissioned iGrids should be two to three times cheaper than imported versions.

Computer software

One of India's major successes in the 1990s is the rapid growth of the computer-software industry, which brought India's high-tech capabilities to the attention of the world. A US\$100-million turnover in 1992 grew to an output of US\$9 billion in 2001, equivalent to 35% of Indian exports and 15% of GDP for that year, with a target of US\$50 billion for 2008.

There have been some accusations that this has been achieved without government support, but the importance of promoting software development, particularly for export, was recognized by the erstwhile Department of Electronics as far back as 1972. Although high trade barriers were then the rule, duty-free import of computers was permitted for companies committing at least 75% of their software output to export. In addition, despite restrictions on direct foreign investment in other sectors, companies that were entirely foreign owned were allowed to set up software export operations in the Santa Cruz export-processing zone on the outskirts of Mumbai. In 1984, the government set up the Software Development and Promotion Agency, but the accelerated growth of the computer industry posed several problems, requiring rationalization of these incentives to promote software exports.

In 1986, software was identified as one of the key sectors on India's agenda for overall export promotion, and emphasis was placed

on the importance of integrated software development for both domestic and export markets. Commercial incentives to software companies included a ten-year tax holiday, 100% income-tax exemption on software-export earnings, export subsidies, and duty-free imports of hardware and software for export purposes. In the 1990s, entry barriers for foreign companies were removed, restrictions on foreign technology transfers lifted, and venture capital was provided to finance software development.

At the same time, free-trade zones for software (software technology parks) were set up with public funds to provide infrastructure for software export. There are now 18 such parks containing 5,600 software export units, which together account for about 70% of India's total software exports.

The government also launched a huge training programme and funded numerous related R&D projects. Without these and other large investments, India's emergence as a major force in the software area just would not have been possible.

Other areas

India has also built up, almost completely independently, a nuclear-energy programme encompassing the entire nuclear fuel cycle, from the mining of uranium ore to the design and manufacture of the equipment for nuclear power stations — as well as developing the expertise needed for installing and commissioning this equipment, and operating and maintaining the stations. India became a nuclear-weapon power in the 1990s, despite numerous restrictions in the form of technology embargoes and sanctions by the existing nuclear-weapon powers.

In addition, India is strong in technologies relating to conventional defence systems, such as electronics systems and satellite technology. This includes a remote sensing satellite with resolution better than 1 metre; a satellite containing telecoms, high-power television broadcasts and meteorological operational payloads; and the Polar Satellite Launch Vehicle, which has not only launched several remote-sensing satellites and multiple satellites in a single launch, but is being prepared for the launch of a 1-tonne scientific experimental payload around the Moon.

There are also, of course, areas in which India has failed to develop and commercialize high technology, notably microelectronics, instrumentation and materials technology. The main challenge for the future is to achieve successes of the kind discussed above on a much larger scale and in many more sectors. Without a doubt, the champion technologies already developed and commercialized must be maintained at such levels and kept internationally competitive. ■

Ashok Parthasarathi is at the Centre for Studies in Science Policy, the School of Social Sciences, Jawaharlal Nehru University, Delhi, India.

Airbrushing science

Scientists are human, so shouldn't the history of science reflect human failings?

Fabulous Science: Fact and Fiction in the History of Scientific Discovery

by John Waller

Oxford University Press: 2002. 320 pp.

£18.99. To appear as *Einstein's Luck in May 2003* (Oxford University Press, \$30).

John Galloway

Science and history make uneasy bedfellows. In *Fabulous Science*, John Waller takes several of our treasured and carefully nurtured illusions about the nature of science and scientists, and systematically uses history to shatter them. Louis Pasteur, Joseph Lister, John Snow, Gregor Mendel — even Charles Darwin — will never be quite the same again.

You might imagine that science and history would lie together rather well. The classical view of history, as expounded by the nineteenth-century German historian Leopold von Ranke, was to show how the past “essentially was”. And what is science if not an attempt to show how nature “essentially is”, how it has evolved, and ideally how our ideas about it have evolved?

But in general, scientists are not interested in history. Eric Larrabee in *Science and the Common Reader* (1966), wrote: “The only thing wrong with scientists is that they don't understand science. They don't know where their own institutions came from, what forces shaped them, and they are wedded to an anti-historical way of thinking which threatens to deter them from ever finding out.” Larrabee's words might well have served as Waller's text.

Let us look at some of Waller's themes. Science has been described as “reliable knowledge”. But if science is what scientists do, then sometimes it is reliable knowledge — or at least it is a process that provides it — and sometimes it isn't. Like other spheres of human endeavour, science divides into the known and the new. There are chaotic, paradigm-shifting ‘wild-west frontiers’; there are areas where knowledge is being consolidated; and there are areas where everything is settled. The first of these is by far the most interesting, but it is also the one where the claims of scientists should be treated most sceptically.

Waller makes this point with some force, taking a variety of examples: Arthur Eddington's (in)famous solar-eclipse experiment to validate the theory of relativity; Robert Millikan's oil droplets; Pasteur's experiments to disprove spontaneous generation and do down his opponents; Frederick Taylor's experiments to provide a scientific basis for management; and Elton Mayo's Hawthorne



Louis Pasteur was lionized in his native France, but used his data to support his own view of the world.

experiments, perhaps designed as a counterblast to them. In all these instances the scientists used their experiments not to test a working hypothesis, but to prove that their point of view was right.

In Raymond Chandler's novel *Playback*, private investigator Philip Marlowe says: “There are things that are facts, in a statistical sense on paper, on a tape recorder in evidence. And there are facts because nothing makes any sense otherwise.” What you see in Waller's examples is the careful selection of facts of Marlowe's first kind to prove that beliefs are ‘facts’ of the second kind. To a committed Catholic creationist like Pasteur, spontaneous generation did not make any more sense than evolution.

And then there is the long-standing tradition of presenting scientists as legendary heroes or secular saints. They were seen as men ahead of their time, whose thinking was resisted by ignorance and vested interests, who were condemned to living in a profes-

sional wilderness, but who ultimately won through in the best Hollywood tradition — acknowledged, lauded and dying with full honours having founded a ‘new’ science.

Waller does a thorough demolition job here. Some heroes, such as the nineteenth-century epidemiologist John Snow, are cut down to a more manageable size, revealed as just one among a number of key players in their particular scientific game. Others, like Joseph Lister and the Charles Best, are revealed as the creators of their own legends, writing their own parts as leading actors and rewriting their own and others' histories as the death or diffidence of their competitors allowed.

Persse McGarrigle, the hero of David Lodge's novel *Small World*, wrote his PhD thesis on the influence of T. S. Elliot on Shakespeare. Defending the thesis against charges that it didn't make sense, he said: “What I try to show is that we can't avoid reading Shakespeare through the lens of Elliot's poetry.”

Waller points out that it seems similarly

impossible for scientists to avoid judging their predecessors through the lens of present-day science. The influence of modern genetics on Mendel in McGarrigle's sense has resulted in a view of Mendel that is completely at variance with the truth, making quite extraordinary claims about him as a man ahead of his time. As Waller convincingly shows, Mendel was actually very much a man of his 'scientific' time in his attempt to unravel the nature of hybridization, an experimental interest he shared with Darwin. The idea that he conducted experiments that changed the world is simply not true. That he was the founding father of genetics is a myth that was generated to aid the social structure of this emerging science.

And finally there is the case of science versus religion. As Waller points out, Richard Dawkins, the arch proponent of evolutionary theory and evangelizing atheist, is in a direct line of ideological descent from Darwin's champion, Thomas Henry Huxley. Dawkins thrives on conflict, real or assumed, between science and religion, just like Huxley, who claimed: "Extinguished theologians lie about the cradle of every science as the strangled snakes beside that of Hercules." The claim was rubbished by John Hedley Brooke in *Science and Religion* (Cambridge University Press, 1991).

But what Waller adds is Huxley's motive: self-interest. Huxley wanted to be a professional scientist but unfortunately there were few paid jobs. Unlike Darwin, who received £5,000 a year from his father, Huxley had to earn his own living. He observed that scientists were often clergymen whose income derived from their parishes. If they could be marginalized scientifically, this might clear the way for himself and others in a similar parlous state to have a paid career. Waller argues that the rift between science and religion was largely one of Huxley's fabrication, with this as his aim.

I have only one slight niggle with *Fabulous Science*. Waller occasionally makes biological mistakes, such as confusing lysosome and lysozyme. Otherwise it is a great read, not just as a historical narrative, but as a commentary, using history to clear away the rubbish that we have been brought up on about the nature of science — by biographers, by the media and, dare I say it, by scientists themselves.

It should now be evident that the 'fabulous' in the book's title is used in its traditional sense as simply the stuff of fables — stories not based on fact that are usually intended to make a moral point. The take-home message is that science is not some intellectual and moral sovereign state. It is a human activity replete with self-deception and promotion, lies, the abuse of power, political manipulation, you name it. On reflection, how could it really be otherwise? ■

John Galloway is at the Eastman Dental Hospital, 256 Gray's Inn Road, London WC1X 8LD, UK.

Our cultural past

Genes, Memes and Human History: Darwinian Archaeology and Cultural Evolution

by Stephen Shennan

Thames & Hudson 2002. 304 pp. £19.95, \$34.95

Robert L. Bettinger

No scientific discipline is so visible to the public and prominent in the popular media, yet so invisible within the sciences proper, as archaeology. Although most archaeologists would prefer the reverse, the situation is partly of our own making. Archaeologists have traditionally imported theory from other disciplines, mainly social anthropology, to the extent that for some scholars 'archaeological theory' is something of an oxymoron.

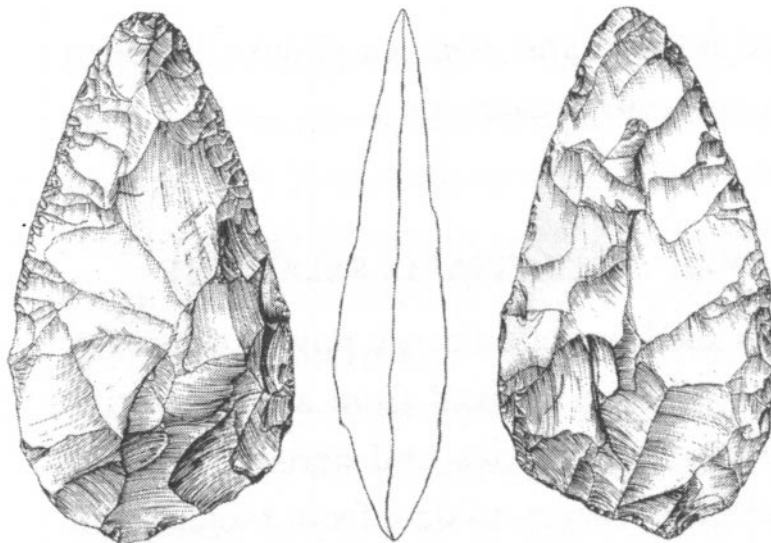
Genes, Memes and Human History accepts that archaeology relies on outside sources for theory — in this case darwinian evolutionary theory, because social anthropology has largely abandoned subject matter that is relevant to archaeology. But the book dispels forcefully the misconception that archaeology is any more dependent in this respect than other fields, such as animal behaviour, in which borrowing carries no stigma. It also demonstrates that modern evolutionary theory makes crucial predictions about processes and behaviours that only archaeology can evaluate. The author, Stephen Shennan, is well qualified to make this case. He is a noted statistician and theorist with a thorough grounding in world archaeology, especially European archaeology, from which many of the book's case studies are drawn.

One of Shennan's main points is that the behaviour of humans, more than that of

other animals, is the product of two analogous, but distinct, systems of inheritance: genetic and cultural. Genetic inheritance is taught in biology courses, and is therefore familiar to most readers, but Shennan reviews the basics, highlighting key implications that may be new to some. Cultural inheritance is less well known as a theory but is intuitively familiar from everyday experience. In cultural inheritance, behaviours are acquired by observing social models.

Parents of teenagers will certainly be familiar with 'conformist transmission', in which individuals simply adopt the most common variant of a behaviour, often from peers (when in Rome, do as the Romans do). There are other forms, too, such as 'guided variation', in which individuals acquire a behaviour from just one social model, perhaps a parent, and modify it through trial and error to suit their individual circumstances and abilities. If others subsequently choose this individual as their social model, the modified trait will be transmitted and may become widespread. Lamarck wrongly supposed that a similar process allows acquired variation in genetically determined traits.

There is a crucial insight here for archaeologists. Because most human behaviour is socially acquired (culturally transmitted), at any given time it is largely a function of historical trajectories that extend over many generations, generally too long for ethnographers to observe, but that are readily detected in the archaeological record. Equally importantly, cultural transmission can produce behaviours that are at odds with the predictions of the genetic model. It is easy to imagine conformist transmission producing a series of locally distinct but internally homogenous groups that set the stage for selection between them. Such group selection is rare elsewhere



Weapon of choice: over-engineering in hand axes may be a sign that males used them to attract females.

in biology, except where groups consist of genetic near-duplicates, such as in eusocial insects. Alternatively, where the social models are the genetic parents, it is easy to imagine guided variation producing behaviours that are essentially in keeping with genetic predictions, because trial-and-error experiment would then be working on parentally inherited variation, as through natural selection.

Shennan uses ethnographic, historic and most of all archaeological evidence to illustrate how these processes are concretely manifested in behaviour and material culture across a wide range of environments and technological settings. For example, prehistoric arrow-points from the Great Basin of western North America are in some places quite variable, suggesting transmission by guided variation using trial and error. In others they are quite uniform, which suggests a conformist-related mode of transmission. These differences have important evolutionary implications; it seems that groups using guided variation perfected adaptive behaviours that allowed them to replace groups whose conformist tendencies prevented them from adopting these innovations.

Humanists sometimes voice the concern that evolutionary analyses of human behaviour are overly simple and deterministic, but Shennan effectively rejects such criticism. As he shows, there is simply too much going on for these critiques to be credible. Genetic and cultural evolutionary processes provide a breathtaking scope of opportunity for the unfolding of interesting, and often conflicting, developments. Cultural transmission may either enhance genetic fitness or compromise it. Cultural traits may be transmitted singly or in bundles that may or may not be functionally related, and at levels ranging all the way from the individual up to the state. These processes occur more or less simultaneously, and the units and levels of analysis must vary accordingly.

Genes, Memes and Human History is eminently readable and deals with quantitative evolutionary concepts in well-written prose, rather than daunting equations. It does not assume substantial knowledge of either archaeology or evolutionary theory, yet does not waste the time of readers who have such knowledge. Archaeologists will find it a helpful source of evolutionary case studies. Scholars in other fields will find it a useful introduction to cultural transmission and to how this and other evolutionary processes are expressed in the archaeological record. Lay readers will find it enlightening in all these respects. It is, in short, a splendid introduction to what evolutionary theory has done, and can do, for archaeology. ■

Robert L. Bettinger is in the Department of Anthropology, University of California, Davis, California 95616, USA.



Skylight: auroras occur when charged particles from the Sun collide with atoms in the atmosphere.

Stormy weather in space

Storms from the Sun: The Emerging Science of Space Weather

by Michael J. Carlowicz and Ramon E. Lopez

Joseph Henry Press: 2002. 256 pp. \$27.95, £19.95

Howard J. Singer

The dictionary definition of storms tells of violent weather, often with examples taken from our everyday experiences of extreme wind and rain. In *Storms from the Sun*, Michael J. Carlowicz and Ramon E. Lopez expand our lexicon of violent weather to include a host of less familiar, but often insidious and significant, space-weather phenomena, such as solar proton events, geomagnetic storms, geomagnetically induced currents and spacecraft charging, to name but a few. They explain both the science and the effects on society of these solar-generated events, which are increasingly being discussed in the media and even in ordinary conversation.

Storms from the Sun is one of a handful of books that bring together the science of the Sun and near-Earth space with the effects of the space environment on human activities and technological systems. The Sun they describe is not the steady, dependable, visible, yellow Sun, but the active and highly variable Sun that emits radiation (such as X-ray, ultraviolet and radio emissions) and energetic particles (mostly protons and electrons). These emissions are invisible without special equipment, but nevertheless

have important consequences for modern technology, resulting in power blackouts, disruptions to communications, difficulty with navigation, problems with satellite operations, radiation hazards to astronauts, and high radiation exposure to passengers and crew on high-flying aircraft.

After an introductory chapter describing a major solar storm in 1998 and its effects on Earth, the authors provide a lucid history of space science from the early records of solar eclipses around 2134 BC to the satellite era and the present day. This history is far more than a timeline of events; it is filled with interesting stories and follows the development of scientific ideas. If you want to know about the connections between Windsor Castle and the aurora, Mark Twain and comets, Columbus and lunar eclipses, *Storms from the Sun* will enlighten and entertain you. A chapter on the solar atmosphere completes the book's foundation.

The remaining chapters draw on relatively recent solar storms as a focal point to describe the science of space weather and its effects on radio communications, satellite operations, astronauts in space, and airline operations. The authors have done an artful job of both educating and entertaining. Explanations of scientific processes are interspersed with historical developments, governmental and industry policy, economic issues and numerous interviews, which add a human touch. The book is therefore likely to appeal both to space scientists and to readers who have a general interest in space science and its impacts on technology and society. Although some passages contain terminology and ideas that may be difficult for non-scientists to grasp, the book's organization makes it easy to skim these passages and move

Science in culture

London's gherkin

The architect Norman Foster has taken 'green' building to new heights.

Michael Hopkin

The latest addition to London's skyline is distinctive for more than aesthetic reasons. The 40-storey, 180-metre-high cigar-shaped office building has also been hailed as a milestone in ecologically friendly architecture. Designed by Norman Foster for the reinsurance firm Swiss Re, the tower has now reached its full height and is due to be completed later this year.

The building is arranged in a series of spirals, each twisting five degrees per storey as they snake upwards. The curved atria or 'lightwells' thus created allow air to circulate throughout the structure — the building's smoothly curving glass skin features opening windows — and the oxygenating indoor gardens that punctuate these atria also prevent the atmosphere from becoming stale. The design cuts air-conditioning costs to a fraction of those in conventional skyscrapers. The lightwells also allow maximum penetration of natural light, reducing the need for electric lighting.

The design is socially friendly to those working in and around the building. Although its shape might seem gratuitous — it has been variously dubbed the 'Erotic Gherkin' and the 'Towering Innuendo' — the aim is to reduce the structure's physical impact on its surroundings. The building's curves will reduce reflected light and prevent winds from being deflected to ground level, as it is by rectangular buildings. And the tower's tapered form allows for more public open space at ground level.

The design has its origins in Foster's collaboration in the early 1970s with Richard Buckminster Fuller, the architect and humanist who lent his name to the football-shaped carbon molecules whose structure resembles his famed domes.

In 1971 the pair came up with a theoretical concept called the 'Climatrotroffice', to feature storey-spanning spaces containing indoor gardens for the circulation and oxygenation of air. Traditional walls and roof are replaced with a 'curvilinear' glass

covering, to minimize building materials. Both of these features are faithfully brought to fruition in the Swiss Re building, which is effectively a high-rise incarnation of the Climatrotroffice.

The idea of glass-skinned buildings was not new even in the 1970s — in 1960 Fuller built a 21-metre-high geodesic dome known as the 'Climatrot' to provide hot and humid conditions for the Missouri Botanical Garden's collection of tropical plants.

Although the Swiss Re tower's environmental credentials are unmatched in London, it is not Foster's first 'green' skyscraper. Europe's tallest building, the 259-metre Commerzbank tower in Frankfurt, which was completed in 1997, also borrows from the Climatrotroffice concept by incorporating a series of nine indoor 'sky gardens'. The gardens — four-storey open spaces — surround a central atrium that pierces the building's entire height. This huge column allows air to circulate throughout the building, providing natural ventilation in place of air conditioning.

Other British architects are also keen to see their country mirror the devotion to ecological considerations shown by continental Europe.



David Marks and Julia Barfield, the architects responsible for the London Eye observation wheel — another of the less conventional shapes on London's skyline — recently unveiled their Skyhouse design as a response to urban planners' demands for affordable housing solutions. The proposed 40-storey towers, which would be composed of petal-shaped wings to make maximum use of natural light, could pair high-density housing with economical energy costs. The contemporary design could also go some way towards subverting the traditional British notion that high-rise living is undesirable.

Ironically, it seems that the world's most skyscraping nation, the United States, is lagging behind in the endeavour to build greener towers. Of the final two proposals competing for New York's Ground Zero site, neither has an ecological focus. Instead, the Lower Manhattan Development Corporation seems determined to make the simple, symbolic statement of erecting the tallest tower in the world. But committing to an environmentally progressive building is a noble response to a terrorist attack — as those in London's financial district are well aware, the Swiss Re building stands on a site whose previous incumbent, the Baltic Exchange, was destroyed by an IRA bomb in 1992.

Michael Hopkin is a subeditor at Nature.



The computer-generated image above shows how the Swiss Re building, with its eye-catching spirals (left), will alter London's skyline.

SWISS RE

on to topics that are more easily absorbed.

The concluding chapters are about the solar cycle, space weather forecasting and thoughts about the future. Appendices with further reading, selected websites, and acronyms and abbreviations will be useful for many readers. The eight pages of colour plates are beautiful and revealing, but are not numbered, and only two plates are referred to in the text.

The book is not without errors, but most of these are minor and are unlikely to be noticed except by experts. For example, at least twice the authors discuss how the solar wind material "tends to fly off in spirals"; whereas its motion is actually known to be

primarily radial, and the spiral appearance that results from changing to a rotating coordinate system could have been explained. Elsewhere the authors have managed to provide simple descriptions of complicated processes.

The authors are even-handed when discussing controversial issues, such as solar influences on climate change; how the United States will capitalize on its investment in research to the benefit of space weather operations; determining how many satellite failures are really due to space weather events or perhaps to other causes; and the apparent reluctance of private companies to talk about space weather problems because of

competitive forces. But most importantly, they bring these issues to the attention of a broader community.

Storms from the Sun covers a remarkable breadth of material that is presented in an appealing format. When asked what they do, professional space scientists now have a book to recommend that explains many aspects of what they study. Policy-makers now have a useful reference book on space weather issues. And non-scientists have an informative and interesting book that explains space science and space weather. ■

Howard J. Singer is in the Research and Development Division, NOAA Space Environment Center, Boulder, Colorado 80305, USA.

Down to Earth

Summarize yourself in the form of a title of a paper in Nature.

Evidence for possible relic biogenic activity in a meteoriticist.

What was your first experiment as a child?

My elder sisters and I performed a series of psychological experiments on my younger sister, but I won't go into details for fear of re-opening old family wounds.

What single scientific paper or talk changed your career path?

When I was an undergraduate, Colin Pillinger gave a talk to my geology class on pre-solar grains and the origin of the Solar System. I knew nothing about planetary sciences before the lecture, and found it amazing that we could learn about things that happened before Earth had formed by looking at tiny grains of meteorites. After the talk I wrote to him and ended up as his PhD student.

What gives you the most job satisfaction now?

Seeing brand new data, and realizing I have found out something that no one knew before.

What are your major frustrations?

Meteorite dealers who understand only the financial value of meteorites, not their scientific worth, and omit or embroider important facts about their find locations and characteristics to make them worth more money. And instruments that are forever breaking down.

What literary character would you employ as a postdoc?

Hermione from the Harry Potter books, because as well as being very clever and studious, she could cast a spell to stop instruments breaking down (see above).

What was the most memorable comment you ever received from a referee?

On a grant proposal: that the work should not be a high priority for funding because the origin of the Solar System was already well understood — in fact, there is an entire research community devoted to this topic.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Dorothy Parker. She would be a fascinating companion and know the best places in New York to spend the evening.

Where and when would you most like to have lived or worked?

In the sample-receiving lab of the Johnson Space Center, Houston, Texas, in 1969 and throughout the 1970s.

What book is currently on your bedside table?

Protostars and Planets IV (used as a lampstand).

What music heads the playlist in your car or lab?

Anything by They Might Be Giants.

What's your favourite conference destination, and why?

The best place would be the Antarctic plateau. It would provide a logistical challenge, but the scenery is stunningly beautiful; the delegates would not get distracted and stray from the conference; and the bracing air would keep everyone wide awake. The 24-hour daylight would mean that it would be possible to hunt for meteorites after the sessions were finished.

What one thing would you rescue from your burning laboratory?

My collection of CAIs (calcium-aluminium-rich inclusions). These are inclusions in meteorites that are among the first solids to have formed in the Solar System. They have many bizarre chemical and isotopic characteristics that are not well understood.

What do you do to relax?

Have a cup of Lapsang Souchong tea.

What would you have become, if not a scientist?

I really like cutting people's hair, when they let me, so hairdressing would be a possibility. Hairdressing is a very creative profession. My grandparents became farmers after they retired, and I might follow in their footsteps as a farmer or a shepherd, although the early mornings might prove too much for me.

What overlooked or underrated discovery really changed the science in which you work?

We now know that there are some grains in meteorites of minerals that formed around stars that were ancestors to our own Sun. These grains can tell us about the chemical evolution of our Galaxy, stellar nucleosynthesis and stellar mixing processes, and about processing of grains in the interstellar medium. Although widely accepted in the meteoritical community, I do not think these discoveries have had as big an impact on the astronomical community as they deserve.

Do you have a burning ambition to do or learn something of no practical or immediate value? If so, what?

To be a good watercolour artist and go on a voyage of exploration as the artist in residence.

Under what conditions do you have your greatest and most inspired ideas?

When I don't have a pen and paper to hand.



Sara Russell

Sara Russell is leader of the meteorites and micrometeorites research programme at the Natural History Museum, London, and lead editor of *The Meteoritical Bulletin*.

What's the one thing about science that you wish the public understood better?

That science and religion are not contradictory. Particularly that the study of the origin of Earth and of life is not anti-religious.

You've just been told (in confidence) that the world will end tomorrow. What do you do next?

Last time this happened, I suggested to the guy who told me that he should recheck his orbital calculations.

Which field in science deserves more funding?

Research into diseases that occur mainly in the developing world, such as malaria and river blindness.

What's the most interesting thing in your fridge?

Chilli vodka, given to me by a friend in Kiev, which is excellent for making Bloody Marys.

Why is physics so hard?

Partly because physicists insist on writing their own computer programs instead of using Windows like normal people.

What would you change about Nature?

I would replace this Lifelines slot with a weekly cosmochemistry piece.

Which actor would best portray you in a film?

Marilyn Monroe.

What would be the title of your autobiography?

'Heavenly Body' (I wish...).

What's just around the corner?

I am looking forward to finding out.

Speciation reversal

Ken Wolfe

Chromosome rearrangements within species are thought to contribute to reproductive isolation between species. This has now been shown directly by unarranging yeast chromosomes to break down a species barrier.

Research into evolution is usually a bit like forensic detective work. Because it's impossible to carry out million-year experiments, we instead look at what evolution has produced and try to figure out what happened and why. Nowhere has this been more difficult than in the study of speciation, the process by which one species splits into two. Biological species are defined by their inability to mate successfully with other species. Usually, we can only speculate about how the barriers to successful mating might have arisen. But on page 68 of this issue¹, Delneri and colleagues describe how they took a hands-on approach, engineering genomes in an attempt to reverse speciation and turn two yeast species into one.

Mating barriers between species can be either prezygotic or postzygotic, depending on whether fertilization can occur (a zygote is the immediate product of fertilization). In the *Saccharomyces* 'sensu stricto' group of yeasts studied by Delneri *et al.*¹, the barriers are postzygotic; different species from this group can mate, but their hybrid offspring are sterile — the fungal equivalent of a mule. Making sterile offspring is a waste of resources, so textbook evolutionary theory suggests that once the fertility of hybrids is low, prezygotic barriers to mating will follow quickly thereafter. The *sensu stricto* yeasts can therefore be regarded as being in the early stages of species differentiation.

How do the first postzygotic barriers between species arise? An attractive hypothesis, termed the chromosomal model of speciation², is that chromosome rearrangements such as reciprocal translocations have a role. A reciprocal translocation is an exchange of material between two different chromosomes, and can simply crop up as a mutation within an individual in a population (Fig. 1a). Mating between individuals with rearranged and unarranged genomes will generate hybrid zygotes. But many of the sex cells (gametes) produced by the hybrids — in yeast, the gametes are spores — will be unviable. This is because they do not contain a complete genome's worth of genes: 25% of the gametes will lack part of one of the translocated chromosomes, and another 25% will lack part of the other chromosome (Fig. 1a). This means that the hybrid strain will be less reproductively successful than its parents, and partial reproductive isolation is established. If additional translocations

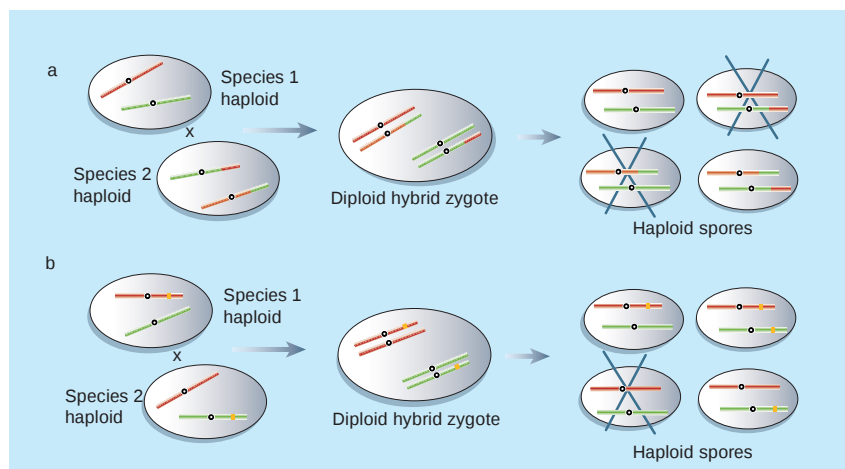


Figure 1 Reduction of hybrid fertility by chromosomal rearrangements. In a, species 1 and 2 differ by a reciprocal translocation that has swapped parts of the red and green chromosomes in species 2. A hybrid formed by mating between these species has reduced fertility, because half of the spores (gametes) that it makes are lacking parts of the genome. In b, one essential gene (yellow) is at a different chromosomal location in species 1 and 2. One-quarter of the spores produced by the hybrid will not survive, because they lack this gene. 'Haploid' and 'diploid' refer to the number of chromosomes present: haploid spores have one copy of each chromosome; diploid zygotes have two.

accumulate, the fertility of the hybrids is reduced even more.

The genomes of two *sensu stricto* yeasts, the *S. cerevisiae* lab strain FY73 and the *S. mikatae* natural isolate IFO1816, are very similar in DNA sequence and are collinear (their genes are in the same order), except that they differ by a reciprocal translocation that swapped parts of chromosomes VI and VII. Is this translocation the reason why hybrids between these yeasts fail to produce viable spores? To find out, Delneri *et al.*¹ modified the genome of FY73, using genetic tools, to create a new strain, Sct1, which has a reciprocal translocation mimicking that in IFO1816. Lo and behold, restoring genome collinearity substantially increased the ability of interspecies hybrids to produce viable spores. Four of the ten hybrids that were tested yielded significant numbers of viable spores (up to 30% of the spores survived), compared with near zero in crosses in which the genomes were not collinear. Further experiments with a second *S. mikatae* strain bearing two reciprocal translocations relative to *S. cerevisiae* FY73, and with hybrids formed between other *sensu stricto* species, demonstrate in a controlled way that genome non-collinearity makes a direct contribution to postzygotic reproductive isolation.

These results suggest that the chromosomal model of speciation is at least partly correct as a mechanism for reinforcing, if not creating, species barriers. What, then, are we to make of the observation that some other yeasts, such as *S. cerevisiae* and *S. paradoxus*, have collinear genomes³ but rarely produce fertile hybrids^{4,5}, and so are regarded as separate species? We first need to recognize that fertility is a quantitative thing, and that in Delneri and colleagues' experiments the interspecies crosses involving collinear genomes produced several hybrids with near-zero spore viability, as well as a few with spore viability increased to 15–30%. This level of fertility is lower than in intraspecies crosses with collinear genomes, but higher than in interspecies crosses involving non-collinear genomes; whether or not it constitutes hybrid sterility is a matter of semantics.

We also need to realize that even when the chromosomes of two species seem to be collinear at a gross level, they can differ in small but highly significant ways, such as the absence of a single gene in one species in an otherwise completely conserved section of chromosome⁶. In the *sensu stricto* yeasts a few exceptions to collinearity of gene order have been documented. Many of these seem to involve genes that were duplicated in an

ancestor, after which the two copies of the duplicated gene diverged in sequence in different offspring, and different copies were ultimately lost⁶. The result is two present-day species with so-called microsyntenic genomic differences, which are detectable only by sequencing. When these species hybridize, the differences act in a similar way to translocations to reduce spore viability⁷ (Fig. 1b). So, the chromosomal model of speciation might prove to be generally true in yeasts, but only when smaller genomic differences, as well as translocations, are taken into account.

There is another intriguing point to the study by Delneri and colleagues. Mating yeasts generally have one copy of each chromosome; the zygotes produced by mating have two copies, one from each parent; and during gamete production (meiosis) the chromosomes are duplicated and segregated at random to produce four spores, each with again just one copy of each chromosome (Fig. 1). But Delneri *et al.* found that, for many of the spores produced by interspecies hybrids, there is a clear tendency to retain the complete set of chromosomes from one parent, plus (as usual) about half the chromosomes from the other. The authors suggest that the chromosome numbers might have been unbalanced in the hybrid zygotes, before they underwent meiosis. This fits with previous findings that hybrid zygotes formed between *Saccharomyces sensu stricto* and the more distantly related *sensu lato* species tend to kick out most of the chromosomes from one of the parents⁸. Making spores that include the complete genome of one parent

could provide an elegant way for a hybrid to overcome the meiotic problems caused by translocations or microsyntenic differences⁴.

Yeasts are proving to be an exciting system for the study of speciation and chromosome evolution, both in nature^{1,4,5} and in artificial culture⁹. The *S. cerevisiae* genome sequence, which represented the mother of all eukaryotic genome projects, will soon be joined by a brood of new genome sequences from closely related yeast species^{10–13}. Because most of these species are amenable to genetic manipulation, the opportunities for experimental dissection of species barriers will be almost limitless. ■

Ken Wolfe is in the Department of Genetics, Smurfit Institute, University of Dublin, Trinity College, Dublin 2, Ireland.
e-mail: khwolfe@tcd.ie

1. Delneri, D. *et al.* *Nature* **422**, 68–72 (2003).
2. White, M. J. D. *Modes of Speciation* (Freeman, San Francisco, 1978).
3. Fischer, G., James, S. A., Roberts, I. N., Oliver, S. G. & Louis, E. J. *Nature* **405**, 451–454 (2000).
4. Greig, D., Borts, R. H., Louis, E. J. & Travisano, M. *Proc. R. Soc. Lond. B* **269**, 1167–1171 (2002).
5. Greig, D., Louis, E. J., Borts, R. H. & Travisano, M. *Science* **298**, 1773–1775 (2002).
6. Fischer, G., Neugebäude, C., Durrens, P., Gaillardin, C. & Dujon, B. *Genome Res.* **11**, 2009–2019 (2001).
7. Lynch, M., O'Hely, M., Walsh, B. & Force, A. *Genetics* **159**, 1789–1804 (2001).
8. Marinoni, G. *et al.* *J. Bacteriol.* **181**, 6488–6496 (1999).
9. Dunham, M. J. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 16144–16149 (2002).
10. Cliften, P. F. *et al.* *Genome Res.* **11**, 1175–1186 (2001).
11. Kamyselis, M., Patterson, N., Birren, B., Berger, B. & Lander, E. S. in *Proceedings of the Seventh Annual International Conference on Computational Biology (RECOMB '03)* (eds Waterman, M. *et al.*) (ACM, New York, in the press).
12. Dietrich, F. S. *et al.* *Curr. Genet.* **35**, 233 (1999).
13. Koszul, R. *et al.* *FEBS Lett.* **534**, 39–48 (2003).

Microwave Anisotropy Probe (WMAP), has returned much higher-resolution images^{2,3} of CMB fluctuations, confirming and extending this remarkable achievement (Fig. 1). These new results provide a worthy book-end to a decade of startling discoveries.

Our Universe is expanding — distant galaxies are moving further away from each other. Using the laws of general relativity and some knowledge of the Universe's constituents, we can trace the evolution of the Universe backwards in time to an era when the density of matter and radiation was significantly higher and a plasma of photons, electrons and nuclei existed in thermal equilibrium. At a temperature of approximately 3,000 K (when the Universe was about 400,000 years old), atomic hydrogen formed and the Universe became transparent. The CMB that we observe has (mostly) streamed freely to us from that moment, cooling as the Universe expanded to a black-body temperature of 2.73 K.

Although the CMB is remarkably uniform from place to place in the sky, small anisotropies in its temperature indicate perturbations in density that are thought to have grown into galaxies and large-scale structure in the contemporary Universe. By assuming that the perturbations originated at very early times, and have roughly equal amplitude at all length scales, we can obtain a good fit to observations of the anisotropies as a function of angular scale. Perturbations of different wavelengths evolve differently as the Universe expands, but they do so in a calculable way that depends on cosmological parameters such as the Hubble constant and the energy density, as well as on the amplitude of the initial fluctuation. Observations of these anisotropies therefore reveal a great deal about the characteristics of our Universe⁴.

Perhaps the most significant aspect of the WMAP results is not the discovery of an unexpected feature of the Universe, but the confirmation of the generally accepted cosmological model that has been constructed over the past several years. In this model, the Universe is spatially flat and 14 billion years old, with an energy density consisting of 30% matter and 70% dark energy (a smoothly distributed component that varies slowly, if at all, as the Universe expands). The matter comes mostly in the form of dark matter, which is believed to be made of a type of particle that is as yet undetected; only 4% of the total energy density of the Universe is ordinary matter (such as stars, planets, gas and dust). Although this model is consistent with a wide variety of observations, it is clearly problematic from various points of view. As ordinary matter and dark matter presumably originate through very different mechanisms, why is their abundance so similar (within an order of magnitude)? Worse still, why is the total abundance of

Cosmology

Filling in the background

Sean Carroll

Data from NASA's Wilkinson Microwave Anisotropy Probe reveal the cosmic microwave background in more detail than ever before. But will cosmology become a victim of its own success?

Not so long ago, cosmology was, half-jokingly, thought of as “a search for two numbers”. The numbers in question were the Hubble constant, which measures the rate at which the Universe is expanding, and the energy density, measured in terms of the critical density for which the Universe is flat (that is, when space has zero curvature). What was worse, many of the field's own practitioners despaired of determining these numbers with any real precision. The difficulty of controlling systematic errors in conventional observations of objects at cosmological distances presented a daunting obstacle.

In the past decade, all this has changed: our description of the Universe is now considerably more detailed, and the most

important cosmological parameters are being determined with precision. The change can be traced back to 1992, when NASA's Cosmic Background Explorer (COBE) satellite first observed¹ small temperature fluctuations, or anisotropies, in the cosmic microwave background (CMB), the relic radiation from the Big Bang. Since then, we have obtained improved measurements of distances to galaxies and supernovae, large-scale structure, light-element abundances, clusters of galaxies, cosmological dynamics and even of the CMB itself. Such measurements have pinned down a precise picture of the state of our Universe that was beyond the hopes of most cosmologists just a short time ago. Now another NASA satellite, the Wilkinson

matter comparable to that of dark energy if they are changing rapidly with respect to each other as the Universe expands? Furthermore, the leading candidate for dark energy is vacuum energy, or the cosmological constant, for which theoretical estimates disagree with observations by 120 orders of magnitude.

But, as ungainly as this model appears, the WMAP results confirm it spectacularly. Previous CMB observations had reached similar conclusions, but it was necessary to stitch together results from many different experiments to cover a wide range of angular scales, and there always lurked the possibility of unknown systematic biases between different data sets. For the first time, WMAP has provided images of CMB anisotropies that cover the entire sky and yet resolve features as small as 0.3 degrees. One obstacle to the construction of all-sky CMB maps is the need to distinguish the sought-after CMB signal from foreground emission from sources both inside and outside our Galaxy. WMAP (like other experiments) achieves this by producing maps in five different frequency bands, and using differences in frequency dependence to separate foreground emission from signal. Fits to the resulting spectrum of anisotropies provide constraints on cosmological parameters that are in excellent agreement with previous data.

But WMAP has done more than simply confirm other experiments. A crucial finding was the measurement of a polarization signal in the CMB, which was first detected by the DASI telescope last year^{5,6}. WMAP has detected polarization at large angular scales, correlated with the temperature fluctuations themselves. This effect, which developed long after the CMB formed, is due to the

reionization of the Universe by high-energy radiation from the first generation of stars. The WMAP findings indicate that these first stars formed when the Universe was only 200 million years old, earlier than previously believed.

A pessimist might worry that the field will become a victim of its own success. The provisional cosmological model fits the observations very well, and one could imagine a future of increasingly precise measurements of cosmological parameters without significant improvements in our fundamental understanding. It seems more likely, however, that a future generation of new experimental approaches (searches for gravitational waves, direct detection of dark matter, measurements of the dark-energy equation of state) will yield their own

surprises. Many of the deepest cosmological mysteries remain — the origin of the matter–antimatter asymmetry, the nature of dark matter and dark energy, and the smoothness of initial conditions in the Universe (perhaps a result of inflation). It's a good bet that the fun is only beginning. ■

Sean Carroll is at the Enrico Fermi Institute, University of Chicago, 5640 S. Ellis Avenue, Chicago, Illinois 60637, USA.

e-mail: carroll@theory.uchicago.edu

1. Smoot, G. *et al.* *Astrophys. J.* **396**, L1–L4 (1992).
2. Bennett, C. L. *et al.* Preprint astro-ph/0302207 at <http://arXiv.org> (2003).
3. http://lambda.gsfc.nasa.gov/product/map/map_bibliography.html
4. Hu, W. & Dodelson, S. *Annu. Rev. Astron. Astrophys.* **40**, 171–216 (2002).
5. Leitch, E. M. *et al.* *Nature* **420**, 763–771 (2002).
6. Kovac, J. M. *et al.* *Nature* **420**, 772–787 (2002).

Immunology

Fast and feel good?

Vijay K. Kuchroo and Lindsay B. Nicholson

Claims that fasting eases symptoms of autoimmune disease have been met with scepticism. But the idea receives some support from the finding that leptin, a hormone that controls body weight, also regulates autoimmunity.

Malnourished people present a tempting target for pathogens. Measles, for example, is renowned for being a more serious disease in populations that are starving or close to starving. Yet many people with the autoimmune diseases rheumatoid arthritis or multiple sclerosis, whose immune systems attack their joints or brain, claim that the symptoms of their disease can be reduced by fasting or a change in diet¹. The idea that food can support a robust immune system and help fight infection, and that starvation can lead to a dampened immune response and alleviate symptoms of autoimmune disease, has mostly been taken as folklore, with little or no scientific basis. But in a paper in the *Journal of Clinical Investigation*, Sanna and colleagues² have shown an intimate relationship between autoimmune disease, starvation and leptin — a key regulator of body weight — in mice.

Leptin is a hormone that was shown in 1994 to be mutated in a certain type of obese mouse³. It is also produced in humans. Leptin is secreted by adipocytes (fat cells) as well as by other tissues, and the initial studies of its function concentrated on its role in regulating food intake and body weight. Recently, leptin has been implicated more generally in insulin metabolism, nutrient use, reproductive function and the response to stress⁴.

Leptin also regulates inflammatory and immune responses and directly affects the immune system, especially T cells⁵. The addition of this hormone to T cells in culture can alter both their growth rate and the

pattern of cytokines — soluble proteins that mediate immune function — that they secrete. For instance, leptin has been shown to enhance the activity of T cells that produce pro-inflammatory cytokines; these same T cells orchestrate many organ-specific autoimmune diseases.

Sanna *et al.*² have now studied the role of leptin during the course of the mouse autoimmune disease experimental autoimmune encephalomyelitis (EAE), which serves as a model of human multiple sclerosis. The authors used several methods to look at leptin's effects. They found that increases in leptin levels in the circulation of mice were correlated with the period before and at the start of clinical disease. They also used mice in which the gene for leptin is non-functional to confirm previous reports that these mice do not develop EAE. Furthermore, T cells that can induce EAE after being transferred into normal mice did not do so after being transferred into leptin-deficient mice. Sanna *et al.* also found that starvation — which reduces leptin levels — inhibited the development of EAE.

This brings us back to the clinical observations that suggested that starvation can regulate symptoms of autoimmune disease. Many controlled trials in humans have shown that fasting and dietary change can ameliorate the symptoms of rheumatoid arthritis, but the explanation put forth — that this is due to a reduction in the levels of anti-food (allergic) antibodies — is difficult to substantiate^{6,7}. In view of Sanna and

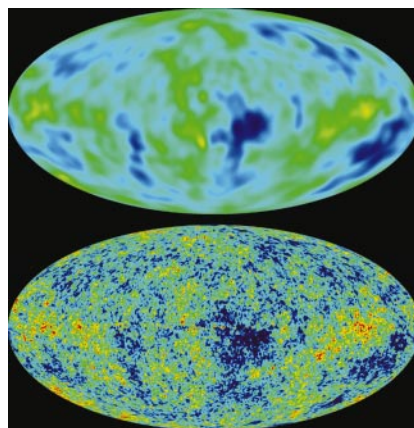


Figure 1 Temperature fluctuations in the cosmic microwave background. In 1992, the Cosmic Background Explorer¹ (COBE) was the first instrument to detect the tiny variations in the temperature of this radiation across the sky (upper image). The latest data from the Wilkinson Microwave Anisotropy Probe^{2,3} resolve the picture in much finer detail (lower image).

colleagues' data, we must also consider whether fasting and changes in diet might be changing leptin levels, thus altering the function of T cells.

The implication of this new work is that leptin drives the activity of pro-inflammatory, self-reactive T cells, and that starvation, which reduces leptin production, changes the pattern of cytokines generated and the disease-inducing potential of the T cells. So, the authors show that instead of producing pro-inflammatory cytokines, T cells from the starved mice generate anti-inflammatory cytokines, which generally inhibit and regulate organ-specific autoimmune diseases. The idea that leptin could also have a significant role in multiple sclerosis is strengthened by studies of the genes expressed in the brains of patients with this disease: leptin and related genes are expressed more than usual⁸.

In the context of the whole animal, however, there is still much to understand about the potential interactions between fat metabolism and immunity. The importance of these questions to multiple sclerosis is highlighted by another recent study⁹ of EAE. This work found that drugs of the statin family, which inhibit the enzyme 3-hydroxy-3-methylglutaryl coenzyme A reductase and reduce cholesterol levels, are also beneficial to animals with EAE, again by altering the cytokine profile of self-reactive T cells. Another enzyme — stearyl coenzyme A desaturase-1, which is involved in fatty-acid synthesis — is thought to contribute to leptin's effects on metabolism¹⁰.

Although these enzymes are on different biosynthetic pathways, their genes are regulated by a common set of gene-transcription factors (adipocyte-determination differentiation-dependent factors, also called sterol-regulatory-element-binding proteins)¹¹ that also regulate the leptin gene¹². So the coordinated control of these different metabolic pathways and leptin could be tied, by transcription factors, both to each other and to aspects of immune function.

In a broader context, these results illustrate once again the trade-off between resistance to infection and susceptibility to autoimmunity. This view is supported by genetic evidence that the regions on chromosomes that influence susceptibility to infection overlap with those that influence susceptibility to autoimmune disease¹³. The demonstration² that starvation can stop EAE reminds us how profoundly our state of immunity is also contingent on our relationship with the environment. Adequate nutrition supports an immune response poised to repulse pathogens. It may also be the best substrate for the seed of autoimmunity to take root.

Vijay K. Kuchroo and Lindsay B. Nicholson are at the Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School,

77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

e-mail: vkuchroo@rics.bwh.harvard.edu

1. Payne, A. J. *Hum. Nutr. Diet.* **14**, 349–357 (2001).
2. Sanna, V. *et al. J. Clin. Invest.* **111**, 241–250 (2003).
3. Zhang, Y. *et al. Nature* **372**, 425–432 (1994).
4. Harris, R. B. S. *Annu. Rev. Nutr.* **20**, 45–75 (2000).
5. Lord, G. M. *et al. Nature* **394**, 897–901 (1998).
6. Kjeldsen-Kragh, J. *et al. Lancet* **338**, 899–902 (1991).

7. Kjeldsen-Kragh, J., Hvatum, M., Haugen, M., Forre, O. & Scott, H. *Clin. Exp. Rheumatol.* **13**, 167–172 (1995).
8. Lock, C. *et al. Nature Med.* **8**, 500–508 (2002).
9. Youssef, S. *et al. Nature* **420**, 78–84 (2002).
10. Cohen, P. *et al. Science* **297**, 240–243 (2002).
11. Edwards, P. A., Tabor, D., Kast, H. R. & Venkateswaran, A. *Biochim. Biophys. Acta* **1529**, 103–113 (2000).
12. Kim, J. B. *et al. J. Clin. Invest.* **101**, 1–9 (1998).
13. Becker, K. G. *Curr. Opin. Allergy Clin. Immunol.* **1**, 399–405 (2001).

Microfluidics

DNA amplification moves on

Andrew J. deMello

The polymerase chain reaction is widely used to amplify samples of DNA for genetic analysis, and fast, high throughput is the ideal. Microscale, chip-based devices are now proving themselves in this arena.

The advent of the polymerase chain reaction (PCR) has hugely accelerated the progress of studies on the genetic structure of a diversity of organisms. PCR is an enzyme-catalysed reaction that allows any nucleic acid sequence to be generated *in vitro*, and in abundance — hence the term 'DNA amplification'. First reported¹ in 1986, PCR has since become a requisite tool in basic molecular biology, genome sequencing, clinical research and evolutionary studies. In an attempt to improve the speed and efficiency of amplification, several groups^{2–4} have created microfabricated systems for PCR, and, in *Analytical Chemistry*, Pierre Obeid and colleagues⁵ report the fabrication of a microfluidic chip that performs both PCR and reverse transcription (the synthesis of DNA using RNA as a template).

The reason for the success of PCR as a DNA-building tool lies in its simplicity. At high temperature (about 95 °C), the double-stranded target DNA denatures — unwinds into two single strands. Synthetic sequences of single-stranded DNA, known as primers, are used to bracket the region of the chain to be amplified: one primer is complementary to one DNA strand (at the start of the target region), with the second primer being complementary to the other DNA strand (at the end of the target region). The primers are annealed to the single strands when the local temperature is reduced to between 50 and 65 °C. This is followed by 'extension', at a slightly higher temperature (about 72 °C), in which a complementary strand develops from each primer by the catalytic action of a DNA polymerase enzyme, in the presence of free deoxynucleoside triphosphates. This three-step process constitutes one PCR cycle, and if repeated *n* times will yield 2^{*n*} copies of the original DNA strand.

Conventional instruments for performing PCR (known as thermal cyclers) are conceptually simple but possess a number of technical frailties that limit the speed and

efficiency of amplification. A fundamental requirement for efficient amplification is rapid heat transfer. It is desirable to have a system with a low heat capacity that can transfer heat quickly to the sample on heating, and quickly away when cooling. Most conventional thermal cyclers have large thermal masses, resulting in high power requirements and slow heating and cooling rates. Consequently, total reaction times are typically in excess of 90 minutes.

As the denaturation and annealing steps occur as soon as the correct temperature is reached, and extension is limited only by the processing power of the polymerase enzyme (between 50 and 100 bases per second), total reaction times can be drastically shortened if the thermal mass of the instrument is reduced. Microfabricated PCR systems have been developed with this idea in mind: although diverse in structure, all rely on the reduction of thermal mass to facilitate rapid heating and cooling, and afford reaction times as short as a few minutes.

The normal approach to instrument miniaturization involves the direct downsizing of system dimensions to reduce thermal masses. An alternative is continuous-flow amplification: instead of heating and cooling a static sample to effect PCR, the sample is moved continuously through multiple reaction zones held at specific temperatures. Originally described by Kopp *et al.*⁶, this approach allows ultrafast reaction times, because the small-volume fluid elements can be heated or cooled to the required temperature within 100 milliseconds.

Obeid *et al.*⁵ have used this concept to demonstrate the functional integration of reverse transcription and PCR within a single microchip. The efficacy of the approach stems from the performance gains afforded through miniaturization and the operational flexibility inherent in the fluidic design. The planar glass chip is fabricated using standard photolithographic and wet-etching techniques and contains a single-channel

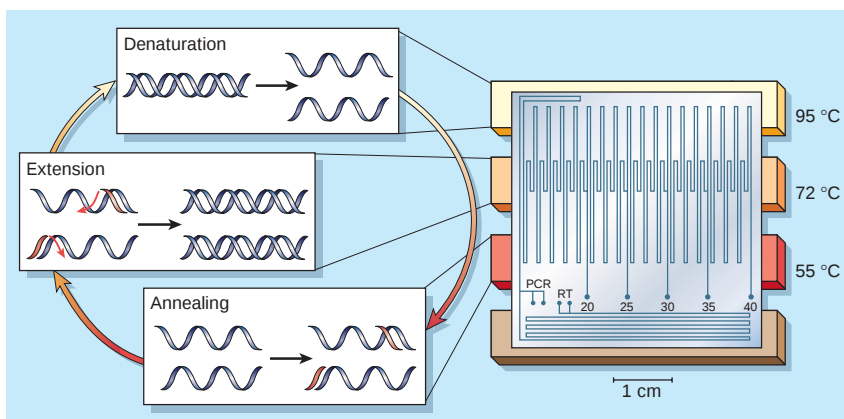


Figure 1 DNA and chips. Obeid *et al.*⁵ have built a microscale device in which DNA can be amplified quickly using the polymerase chain reaction (PCR). When a sample is added at either of the PCR inputs it flows over heating blocks whose temperatures are set to induce the three steps of PCR consecutively: denaturation, or unwinding of the DNA strands; annealing, in which primers are attached to the separated strands; and finally extension of the primers into complete DNA strands, to produce two copies of the original DNA strand. Samples can be extracted from the chip after between 20 and 40 PCR cycles, at the points indicated. A further advantage of this design is the channel for reverse transcription (RT), in which RNA samples can be transcribed into DNA before entering the PCR region of the chip for amplification.

network defining four zones, one for reverse transcription and three for PCR (Fig. 1). Zone temperatures are controlled by simply placing the entire chip over four temperature-controlled heating blocks. Five outlets for product collection are located along the channel, so the product can be analysed after 20, 25, 30, 35 and 40 PCR cycles. Using their device, the authors demonstrate efficient amplification of DNA after 20 cycles in times as short as 5 minutes. Furthermore, by

periodically injecting small samples (2 μ l) into the system, separated by water and air plugs, simultaneous amplification of multiple samples can be performed in continuous flow without cross-contamination.

The benefit of continuous-flow amplification is further demonstrated by this device's ability to perform reverse transcription of RNA into DNA before PCR amplification—a process widely used for the quantification of messenger RNA levels.

Reverse transcription is performed within a serpentine microchannel that, downstream, intersects the PCR channel, and subsequently proceeds through the heating zones. Integration of the two processes within a monolithic device is often problematic as reverse-transcription components at high concentration can interfere with the PCR. The authors tackle this problem by reducing the flow rate at which reverse transcription is performed, so that, at the intersection of the reverse-transcription and PCR channels, the reverse-transcription mixture constitutes about 10% of the PCR volume. With this approach, high-throughput reverse-transcription-PCR (of 0.7- μ l volumes) is achieved in short times and without nonspecific amplification.

The work described by Obeid and colleagues demonstrates the true integration of biologically relevant processes within a monolithic device. Importantly, continuous-flow operation offers a direct route to automated sample introduction, mixing and reaction, and thus the possibility of high-throughput sequence analysis in many practical applications.

Andrew J. deMello is in the Department of Chemistry, Imperial College, London SW7 2AZ, UK.
e-mail: a.demello@imperial.ac.uk

1. Mullis, K. B. *et al.* *Cold Spring Harb. Symp. Quant. Biol.* **51**, 263–273 (1986).
2. Woolley, A. T. *et al.* *Anal. Chem.* **68**, 4081–4086 (1996).
3. Northrup, M. A. *et al.* *Anal. Chem.* **70**, 918–922 (1998).
4. Lagally, E. T., Medintz, I. & Mathies, R. A. *Anal. Chem.* **73**, 565–570 (2001).
5. Obeid, P. J., Christopoulos, T. K., Crabtree, H. J. & Backhouse, C. J. *Anal. Chem.* **75**, 288–295 (2003).
6. Kopp, M. U., deMello, A. J. & Manz, A. *Science* **280**, 1046–1048 (1998).

Climate change

The earlier bird

Since 1909, researchers have been catching and marking migrating birds that stop over on the island of Helgoland in the southeastern North Sea. These birds breed in Scandinavia and spend the winter in either continental Europe (short-distance migrants) or Africa (long-distance migrants). The methods of trapping — one type of apparatus is shown in the photograph opposite — have not changed since 1960. Moreover, the data cover around two dozen species, and describe the mean time of migration for all trapped individuals, not just extremes in the form of first arrivals. All of this makes the Helgoland data sets some of the best available with which to study the timing of bird migration.

Ommo and Kathrin Hüppop have now analysed these data sets, with

remarkable results (*Proc. R. Soc. Lond. B* **270**, 233–240; 2003). They find that all 23 migratory bird species for which sufficient data are available pass by Helgoland on their way to Scandinavia earlier now — by two to twelve days — than 40 years ago. There is a clear division between short-distance migrants, whose mean time of passing correlates well with local temperatures, and long-distance migrants, for which increases in the NAO index (a measure of the air pressure over the North Atlantic Ocean) give a much better explanation for the earlier time of passage.

Whether a migrating bird actually lands on a small island while passing over it depends on many factors. Sudden changes in weather play a large part. So it is important to analyse large data

sets to disentangle general patterns from such isolated examples. The changes in the timing of migration are apparently strong enough to become evident.

These changes, in particular the earlier passage of the long-distance migrants, raise questions about the control of spring migration. There are three main hypotheses that might explain their earlier passage. First, the moment of leaving Africa has not changed, but refuelling in continental Europe proceeds more quickly, because more food — in the form of insects — is available earlier. (Increases in the NAO index generally indicate favourable spring conditions in Europe.) Second, if the weather in Africa is also correlated with the NAO index, then the birds might leave earlier because the seasons there also change earlier. Third, the weather in



Africa has not changed, but natural selection has altered the 'trigger values' for starting migration. Each hypothesis, if true, would mark an exciting break with existing knowledge, and I eagerly await further results on changes in spring migration.

Arie J. van Noordwijk
Arie J. van Noordwijk is at the Netherlands Institute of Ecology, Boterhoeksestraat 48, 6666 GA Heteren, The Netherlands.
e-mail: a.vannoordwijk@nioo.knaw.nl

Planetary science

Kuiper-belt interlopers

A. Morbidelli and H. F. Levison

Objects in the Kuiper belt, which lies beyond Neptune, occur as two distinct populations. One group may have migrated from a region closer to the Sun, caught by Neptune's gravity as it wandered in the early Solar System.

When astronomers first postulated the existence of a belt of small bodies beyond Neptune — now called the Kuiper belt — they imagined a disk of planetesimals (primordial icy bodies) in which the conditions of the early Solar System were perfectly preserved. But when these objects were finally discovered, astronomers realized that this is not so: the Kuiper belt is dynamically excited, as if something had kicked it very hard. The origin of this excitation is a central question in planetary astronomy today, because it could give vital clues to the origin of the planets. Writing in *Icarus*, Rodney Gomes¹ proposes a new mechanism that can explain a lot of what we see. If he is right, much of the structure of the Kuiper belt has been generated by a slow, outward migration of Neptune's orbit during the last phases of its birth.

A detailed analysis² of the orbits of objects in the Kuiper belt — and one that accounts for observational biases — shows that the vast majority of the bodies can be divided into two populations of roughly equal numbers: a 'cold' population whose orbits are inclined at only a few degrees to the average orbital plane of the planets (called the invariant plane); and a 'hot' population that has a broad distribution of orbital

inclinations, extending up to 40°. The terms hot and cold do not refer to temperature, but rather to the level of dynamical excitation of the bodies (as seen in the inclinations and eccentricities of their orbits). The co-existence of these two populations in the Kuiper belt has puzzled astronomers.

Even more puzzling is the growing evidence that the bodies in the two populations are physically different. In the absence of detailed spectra, astronomers classify objects according to the colour of the light that they reflect from their surface. The colour distributions in the cold and hot populations are statistically different: the hot population has predominantly grey colours, whereas the cold population is almost exclusively made up of red objects^{3,4}. In addition, all the largest Kuiper-belt objects seem to belong to the hot population⁴, although the evidence for this is statistically weak.

Through simulations of planetary migration and the evolution of planetesimal orbits, Gomes has devised a mechanism¹ that explains the origin of the hot population and also implicitly accounts for the differences in its physical properties with respect to the cold population. The starting point is the migration of Uranus and Neptune. After their formation, these planets are generally

believed to have moved further away from the Sun, because of the gravitational effect of the scattering of planetesimals in their vicinity^{5,6}. Under the gravitational influence of Neptune, and to a lesser extent of Uranus, most of these planetesimals were transported closer to the Sun before being ejected from the Solar System through encounters with Jupiter or Saturn (Fig. 1). As energy and angular momentum had to be conserved, and as Neptune had sent most of the material inwards, that planet slowly spiralled outwards as a result.

However, some planetesimals were transported outwards by Neptune, on elliptical orbits. These orbits must have regularly crossed that of Neptune, and at each encounter the planetesimals would have been knocked into new orbits, at new mean distances from Neptune and new inclinations to the invariant plane. Occasionally, an object would have entered a resonance with the planet. Resonances are subtle dynamical phenomena that occur when, for example, the orbital period of the body is in a simple integer ratio with that of the planet. In a resonance, the ellipticity of the orbit can be modified without a planetary encounter.

If the ellipticity of a planetesimal's orbit decreases (becoming more circular), the sequence of encounters eventually stops and the planetesimal becomes 'decoupled' from Neptune, like a classical Kuiper-belt object. This process is reversible: eventually, the ellipticity increases again to the extent that the orbits of Neptune and the planetesimal again cross, and gravitational encounters restart. But Gomes has found that Neptune's migration away from the Sun broke this reversibility: as a result, some of the decoupled bodies escaped from their resonances and remained permanently trapped in the Kuiper belt.

Thus, the Kuiper belt is inhabited by two populations: the local population, which probably formed *in situ* and, never having had encounters with Neptune, essentially preserved its small-inclination distribution; and the population of bodies that immigrated from the region around Neptune, with large inclinations acquired during their Neptune-encountering phase. These two populations can be identified respectively with the cold and hot populations that are observed. Indeed, in Gomes' simulations, the inclination distribution of the immigrants matches that observed in the hot population very well. And, having originated in two different regions of the Solar System, the two populations naturally have different physical properties. The immigrant population would also be expected to host larger objects than the cold population: the accretion rate (at which planetesimals stick together to form larger bodies) would have been higher at smaller distances from the Sun.

Ten years after the discovery of the first

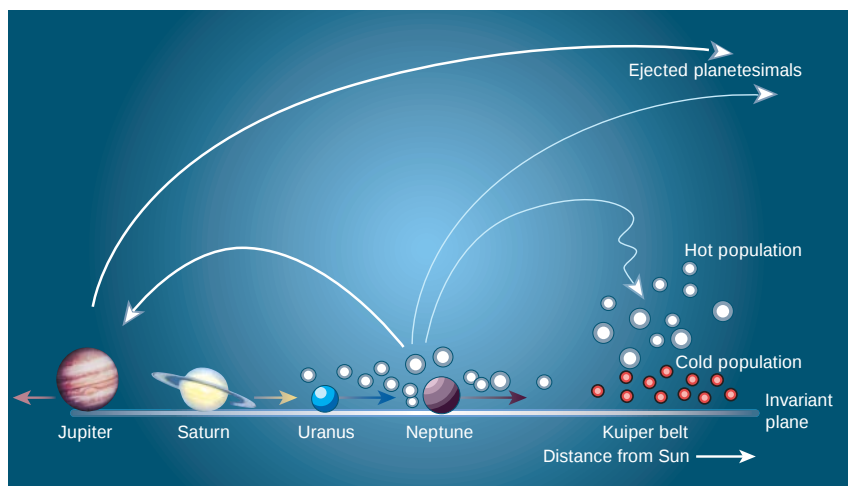


Figure 1 On the origin of the 'hot' Kuiper-belt population. Gomes¹ proposes that this group of planetesimals, now in orbit beyond Neptune, originated in a region between the current orbits of Uranus and Neptune. Gravitational effects in the developing Solar System threw some of these primordial bodies out of the system completely; others were deflected towards Jupiter and Saturn before they too were ejected. But Neptune's migration away from the Sun (balancing the inward motion of planetesimals towards Jupiter and Saturn) trapped some objects that would otherwise have been lost. This mechanism explains why the properties of the hot population are different from those of the cold population that evolved *in situ* and is confined at small inclinations to the invariant plane.

true Kuiper-belt object, Gomes' work gives a new perspective on these bodies. If the hot Kuiper belt is really a record of the Uranus–Neptune planetesimal population, a comparison of the detailed physical properties of the hot and cold objects could provide valuable information on the way in which the temperature and chemical properties of bodies varied with increasing distance from the Sun in primordial times. As astronomers slowly unlock these, and other, mysteries, we will learn more about the early life of our planetary system. ■

A. Morbidelli is at the Observatoire de la Côte d'Azur, 06304 Nice, France.
e-mail: morby@obs-nice.fr

H. F. Levison is at the Southwest Research Institute, 1050 Walnut Street, Boulder, Colorado 80302, USA.
e-mail: hal@boulder.swri.edu

1. Gomes, R. S. *Icarus* **161**, 404–418 (2003).
2. Brown, M. *Astron. J.* **121**, 2804–2814 (2001).
3. Trujillo, C. A. & Brown, M. E. *Astrophys. J.* **566**, 125–128 (2002).
4. Levison, H. F. & Stern, S. A. *Astron. J.* **121**, 1730–1735 (2001).
5. Fernandez, J. A. & Ip, W. H. *Planet. Space Sci.* **44**, 431–439 (1996).
6. Hahn, J. M. & Malhotra, R. *Astron. J.* **117**, 3041–3053 (1999).

Plant biology

Mobile plastid genes

Pal Maliga

The direct demonstration that chloroplast DNA can be incorporated into the nuclear genome of plants, even though it is unlikely that such DNA would be functional, will influence thinking in plant biotechnology.

Plastids are organelles found in the cells of higher plants, the best-known form being the chloroplast, the site of photosynthesis. They arose when, way back in time, plant ancestors assimilated photosynthetic prokaryotes — unicellular organisms such as bacteria which, in contrast to eukaryotes such as plants (and ourselves), typically lack a membrane-bound nucleus. The creation of functional plastids involved the migration of huge numbers of genes from the prokaryote genome to the plant nucleus. But is transfer of plastid DNA into the nucleus of higher plants still occurring?

As they describe on page 72 of this issue¹, Timmis and colleagues conclude that it is, at a rate comparable to spontaneous mutation of the nuclear DNA. There are various implications here, not least for plant biotechnology. One proposed strategy for the future is to engineer genes for some trait or other into chloroplasts, rather than nuclei, because — in principle — those genes will affect the characteristics of the individual plant and will be transmitted to future generations by the maternal parent, but not 'escape' in pollen and spread uncontrollably.

The extent of gene migration during evolution becomes clear when we consider that the plastids of higher plants today contain only about 120 genes, compared to the 3,000–4,000 thought to have been present in the genome of the ancient photosynthetic prokaryote. Plastid function depends on some 3,000 nuclear genes, the products of which are transferred to complement those of the remaining plastid genes². We have a relatively good idea about the identity and gene content of the organisms that were involved when the original eukaryotic plant ancestor took in the prokaryotic forerunner of the plastid³. But little is known about the

evolutionary process of gene transfer from the plastid to the nucleus.

This is what Timmis and colleagues¹ set out to study by designing a smart screen to estimate the rate of plastid-to-nucleus DNA transfer in tobacco plants. Their system involved measuring the transfer rate to the plant nucleus of a gene (*neo*) — which confers resistance to the antibiotic kanamycin — that had been engineered into the tobacco plastid genome (along with a marker gene, *aadA*, of which more later). The kanamycin-resistance gene was endowed with the various signals, including a promoter sequence, required for eukaryote-type expression in the nucleus. Plastids, in general, have a prokaryotic-type gene-expression machinery⁴, so the nuclear kanamycin-resistance gene was not expected to be expressed there (Fig. 1a, overleaf).

Timmis's group found that transfer of *neo* to the nucleus took place in a total of 16 heritable events in 250,000 seedlings — that is, at an incidence of around 6×10^{-5} — and was detected by testing for kanamycin resistance in the tobacco plants (Fig. 1b, c). Molecular analyses confirmed that the *neo* genes, together with flanking plastid DNA of variable size, had indeed been incorporated into the tobacco nuclear DNA at different genomic locations.

Another cellular organelle found in eukaryotes is the energy-generating mitochondrion. Previous work⁵ with yeast has shown that here, too, there is transfer of organelle DNA to the nucleus, and the rate (about 2×10^{-5} per cell per generation) is comparable to that found by Timmis *et al.* in tobacco. In yeast, the fragments of mitochondrial DNA were incorporated into the chromosomes by a mechanism known as 'double-strand-break repair'^{6,7}, which is



100 YEARS AGO

All inquirers have perceived that great men are of two types, and it would conduce to clear thinking if we could accustom ourselves to classify them under different names... The first class, to which I should prefer to restrict the name genius, may be described primarily as men of fine, delicate, sensitive, impressionable constitution, and strong, restless innate tendencies which appear early in life, as a rule, and take their own shape. These men work energetically, often at high pressure, and in general die comparatively young... The second class I would describe as men of talent. When preeminent they exhibit striking aptitude in learning and in imitation, and develop extraordinary powers of work... In nature there is great variety, and genius, so far, is one of the varieties which often recur, but scarcely ever survive even for two generations. It is a rare and delicate thing, and the utmost we can hope for it is that endeavours may be made to collect and preserve it like some hot-house plant, in order that it may suggest combinations which men of talent may put to practical account. The position of the second type in the struggle for existence is beyond doubt. The stability of a country and its place among the nations depend upon the number and ability of men of this stamp.

From *Nature* 5 March 1903.

50 YEARS AGO

There has been a progressive increase, both absolute and relative, in the proportion of old people in the population, this increase being the result of the decline in infant and young adult mortality produced by the medical and social advances that have been made possible by the application of the scientific method during the past century. As a result of this, more old people must be supported by a diminished proportion of wage earners. There is a corresponding increase in the numbers of elderly invalids to be cared for, who are suffering from the degenerative disease of old age which medical treatment may ameliorate, but not cure. Meanwhile, the expense of medical treatment, especially hospital treatment, has risen enormously and continues to rise, and it has been suggested that the community is faced with a gloomy prospect of unlimited 'medicated survival' to be met by diminishing resources.

From *Nature* 7 March 1953.

usually involved in repairing damaged DNA. The same mechanism might be involved in the transfer of *neo*-containing plastid DNA, although an earlier study⁹ of double-strand-break repair in tobacco found no evidence of plastid (or mitochondrial) DNA transfer during the process.

Given the readily detectable plastid-to-nucleus transfer of DNA, it is not surprising that the tobacco nuclear genome contains long tracts of plastid DNA. For example, a search⁹ for plastid DNA fragments containing a particular gene, *rbcL*, found that there are 15 fragments of different size containing *rbcL* plastid DNA in the tobacco nuclear genome, the largest being 15.5 kilobases (kb) in size. But the amount of organellar DNA present in the nucleus seems to depend on the plant species concerned. For instance, the nuclear genome of *Arabidopsis thaliana*, the favoured model for plant biologists, contains only 17 small insertions of plastid DNA (total size 11 kb)¹⁰, and 13 small insertions (7 kb)¹⁰ and one large insertion (about 620 kb) of mitochondrial DNA¹¹. The differences may in part be due to different organelle-to-nucleus DNA transfer rates. Indeed, 40% of double-strand-break repair in the tobacco nucleus was accompanied by insertions (although, in the specific instance, not of organellar DNA), whereas no insertions accompanied double-strand-break repair in *Arabidopsis*⁸. If the conclusions of the study by Timmis *et al.*¹ can be generalized, crops that have larger nuclear genomes are likely to have higher transfer rates between plastids and the nucleus. But the relative abundance of organellar DNA in nuclear genomes may also reflect the efficiency of mechanisms that work against 'genome obesity' by selectively eliminating nuclear sequences¹².

What about the biotechnological implications of the new research? In general, incorporation of genes for some trait or other into the plastid genome is an attractive strategy, as plastids in most crops are not transmitted by pollen. So incorporation of, for instance, genes conferring herbicide or insect resistance in the plastid genome would mean that these genes would not be transmitted by pollen to neighbouring stands of the same crop or to weedy relatives¹³. This is a common problem with transgenes incorporated in the nuclear genome¹⁴.

It is reassuring that the marker gene (*aadA*, which confers resistance to another antibiotic, spectinomycin) that Timmis and colleagues incorporated into plastids along with the kanamycin-resistance gene, was not expressed in the nucleus: this gene had a plastid-specific promoter. But although *aadA* residing in the nucleus on a plastid genome segment is not expressed, broken plastid gene pieces, incorporated next to a nuclear promoter, could be expressed at a very low frequency¹⁵.

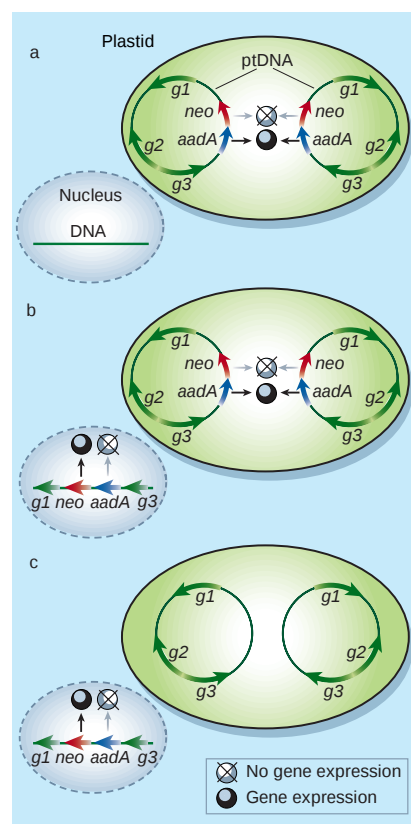
The authors tested for spectinomycin resistance only in the 16 plant lines carrying

Figure 1 **Transfer of plastid DNA to the nucleus of a tobacco cell, as studied by Timmis and colleagues¹.** a, Tobacco cell with a transformed plastid genome (ptDNA) and a wild-type, non-transformed nucleus. Three plastid genes (*g1*, *g2*, *g3*) are shown in the plastid genome, along with the kanamycin-resistance (*neo*) gene. Crucially, *neo* has the necessary ancillary sequences for expression only in the nucleus, whereas *aadA* can be expressed only in the plastid. Only two ptDNA copies are shown out of the 1,000–10,000 copies present in a plant cell. b, A rare cell, in which a fragment of ptDNA has been transferred to the nucleus: note that *neo* is expressed here but *aadA* is not. c, Timmis *et al.* had to study the expression of plastid DNA in the nucleus in cells with wild-type ptDNA, the cells being created by pollinating non-transformed plants with pollen from plants possessing transformed plastids. One in 16,000 pollen grains carried a *neo* gene inserted in a chromosome. Because plastids are not transmitted by pollen, only the nuclear *neo* gene (and flanking ptDNA) was transmitted in the cross. Separation of plastid and nuclear transgene copies was essential as there are many more plastid *neo* and *aadA* copies than nuclear copies, and the large number (10,000 per leaf cell) of plastid copies interferes with the study of the few copies in the nucleus.

the transferred *neo* gene. To estimate the transfer rate of functional *aadA*, however, a seedling population of comparable size (250,000) or larger would have to be screened for expression of spectinomycin resistance. The frequency of fortuitous transfer that results in expression of the transferred plastid gene may be 100–100,000 times lower than the rate of transfer of the nuclear *neo* gene (the values will have to be experimentally determined).

Nonetheless, incorporation of transgenes in plastids should still be effective for containment of those genes. If a transgene is incorporated in the nucleus, each pollen grain will carry the transgenic trait. By contrast, assuming that the probability of plastid gene expression from a broken fragment is only 100 times lower than the transfer rate of plastid DNA, only 1 in 1.6 million pollen grains will carry the expressed plastid gene.

It could be, however, that functioning of the transferred plastid genes in the nucleus could be prevented altogether. Examples of genes that are expressed in plastids but not in the nucleus are bacterial genes containing a high content of adenine and thymine, such as those encoding the *Bacillus thuringiensis* insecticidal protein or the *Clostridium tetani* toxin^{13,16}. The *B. thuringiensis* insecticidal genes could be expressed in the plant nucleus only by modifying them to a 'eukaryotic design'^{17,18}. So modifying a gene to a 'bacterial design' would be expected to facilitate its expression in plastids and prevent it in the nucleus.



Timmis and colleagues' demonstration¹ of the transfer of plastid DNA to the nucleus of higher plants will be the subject of much debate from both the scientific and biotechnological viewpoints. The implications in the latter respect will depend on the likelihood of plastid genes becoming functional nuclear genes, and on the success of gene designs that can prevent the expression of transferred plastid genes in the nucleus. ■

Pal Maliga is at the Waksman Institute, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854-8020, USA.

e-mail: maliga@waksman.rutgers.edu

- Huang, C. Y., Ayliffe, M. A. & Timmis, J. N. *Nature* **422**, 72–76 (2003).
- Abdallah, F., Salamini, F. & Leister, D. *Trends Plant Sci.* **5**, 141–142 (2000).
- Palmer, J. D. *Nature* **405**, 32–33 (2000).
- Sugiura, M. *Plant Mol. Biol.* **19**, 149–168 (1992).
- Thorsness, P. E. & Fox, T. D. *Nature* **346**, 376–379 (1990).
- Ricchetti, M., Fairhead, C. & Dujon, B. *Nature* **402**, 96–100 (1999).
- Yu, X. & Gabriel, A. *Mol. Cell* **4**, 873–881 (1999).
- Kirik, A., Salomon, S. & Puchta, H. *EMBO J.* **19**, 5562–5566 (2000).
- Ayliffe, A. M. & Timmis, J. N. *Theor. Appl. Genet.* **85**, 229–238 (1992).
- The Arabidopsis Genome Initiative. *Nature* **408**, 796–815 (2000).
- Stupar, R. M. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 5099–5103 (2001).
- Bennetzen, J. L. *Genetica* **115**, 29–36 (2002).
- Maliga, P. *Trends Biotechnol.* **21**, 20–28 (2003).
- Riegel, M. A. *et al.* *Science* **296**, 2386–2388 (2002).
- Carrer, H., Hockenberry, T. N., Svab, Z. & Maliga, P. *Mol. Gen. Genet.* **241**, 49–56 (1993).
- Tregoning, J. *et al.* *Nucleic Acids Res.* **31**, 1174–1179 (2003).
- Perlak, F. J., Fuchs, R. L., Dean, D. A., McPherson, S. L. & Fischhoff, D. *Proc. Natl Acad. Sci. USA* **88**, 3324–3328 (1991).
- Murray, E. E. *et al.* *Plant Mol. Biol.* **16**, 1035–1050 (1991).

Immunology

Geared up for antibody production

Immunity **18**, 243–253 (2003)

When the B cells of our immune system detect unfamiliar items, they turn into plasma cells that are dedicated to producing and secreting antibodies that cleanse our body. Eelco van Anken *et al.* find that B cells meticulously plan this transformation.

Antibody production is an elaborate process. First, individual subunits are synthesized and dispatched to the endoplasmic reticulum. Inside this network of membranes, chaperone proteins help the subunits to fold into shape and assemble together. The antibodies are then transported to the cell surface and released — up to thousands per second.

Van Anken *et al.* used a dynamic proteomics approach to show that sequential waves of functionally related proteins are synthesized during B-cell transformation. For instance, the cells first stock up on mitochondrial and cytosolic chaperones, probably to sustain increased energy consumption and protein synthesis. Supplies of metabolic enzymes are boosted later, perhaps reflecting the need for expansion of the endoplasmic reticulum membrane. Meanwhile, the numbers of chaperones in the endoplasmic reticulum build up steadily. The production of antibodies increases exponentially, but only after a lag period — allowing the cell to build an efficient antibody production plant first.

Marie-Thérèse Heemels

Chemistry

Levitation lights up crystals

Anal. Chem. doi:10.1021/ac020496y (2003)

Growing protein crystals for structure determination is a complex process, often achieved only through trial and error. But Sabina Santesson and colleagues hope, through filming protein precipitation in levitating liquid drops, to better understand and control crystal growth.

The team levitate a microlitre drop of supersaturated protein solution by trapping it in a standing acoustic wave. They then inject crystallization agents or water into the drop. Measurements of light-scattering indicate the onset of precipitation, while microscope imaging records the drop's volume and hence the concentration of its components.

With this set-up, minute quantities of protein are sufficient to rapidly screen different crystallizing agents and map out the limits beyond which supersaturated solutions become unstable and the protein

precipitates. This information guides the choice of conditions for more detailed batch nucleation tests.

Santesson *et al.* show that their approach readily yields crystals of the binary complex formed between the enzyme D-serine dehydratase and pyridoxal-5'-phosphate. A crystallization screening kit apparently gave poorer results.

Magdalena Helmer

Neurobiology

Quirks of memory

Curr. Biol. **13**, 286–296 (2003)

Quirky names are often given to mutant fruitflies. Even so, Tim Tully and his colleagues seem to have taken the off-beat practice to extremes by calling a fly with a memory problem *krasavietz* — Russian for 'male beauty'. History has informed their choice, however, for Krasavietz was one of Ivan Pavlov's dogs, with which the Russian physiologist established associative conditioning as a paradigm for learning and memory.

Tully and co-workers screened flies by using behavioural tests to identify mutants that had defects in their long-term memory (naming them all after Pavlov's dogs), and also carried out DNA-microarray analysis to identify the gene-transcription machinery associated with such memories. This impressive two-pronged attack enabled the authors to identify components of the localization and translational apparatus of the messenger RNAs involved: the *krasavietz* mutant, for instance, corresponds to damage to a translation-initiation factor known as *eIF-C*. In one follow-up example, Tully *et al.* used temperature-sensitive mutants of *staufen*, which influences the translocation of mRNA, to study the critical timing of this gene's expression in the formation of long-term memory.

Hemai Parthasarathy

Astronomy

The long and the short of it

Astron. Astrophys. (in the press); Preprint astro-ph/0301262 at <http://arXiv.org> (2003)

Depending on whether they are short or long, mysterious γ -ray bursts (GRBs) probably originate from different cosmic sources. GRBs are the most powerful explosions in the known Universe, presaging the formation of black holes. They may last only a few milliseconds, or up to 100 seconds; most last only about 10 seconds.

Previous work had hinted at separate physical sources for short and long bursts, a suspicion now reinforced by the results of L. G. Balázs and colleagues. They have analysed

the light signatures of 1,972 GRBs, detected using NASA's Compton Gamma Ray Observatory. Balázs *et al.* find that for bursts lasting longer than two seconds the flow and energy — termed 'fluence' — of the γ -rays is directly proportional to the burst's duration. But for short bursts this relationship is less clear, suggesting perhaps that for these GRBs the source converts energy into γ -rays at a rate that falls with time.

Long bursts might be caused by the explosion of massive stars (more than 30 times heavier than our Sun), and short bursts by the merging of ultra-dense neutron stars. But improved models of GRBs are needed before these origins can be confirmed.

Tom Clarke

Evolutionary biology

Species, more or less

J. Evol. Biol. **16**, 282–288 (2003)

Some groups of animals contain many more species than others. M. Cardillo and colleagues have investigated the reasons for this, taking Australian mammals as their study subjects. They conclude that, in this case at least, genera with many species seem to have features in their ecology that buffer them against extinction.

Cardillo *et al.* found that the distribution of species among genera can't be explained by a random distribution, or by the oldest groups containing the most species. Genera with many species, therefore, could possess traits that promote the formation of new species or help to avoid extinction (or both).

It turns out that the number of species in a genus correlates most closely with the geographical size of that group's range and with its litter size. The authors suggest that both of these characteristics make it less likely that a species will become extinct. A large range gives a widespread population that can recolonize particular sites from which a species has disappeared, and species with large litters have fast rates of population growth and so can bounce back from knocks. Body size seems to make no difference — a surprise, as larger animals seem to have suffered higher extinction rates over the past few millennia. Human hunting may be the explanation there.

John Whitfield



Disease susceptibility in California sea lions

Inbreeding influences the response of these animals to different pathogens in the wild.

Inbreeding in animals can increase their susceptibility to pathogens, but direct evidence from wild populations is scarce^{1,2} and it is unclear whether all pathogens are affected equally. Here we analyse rescued California sea lions afflicted with a range of different pathogens, and find that sick animals have higher-than-normal parental relatedness, with the extent varying among disease classes. Our findings indicate that mortality in natural populations may not be entirely random and that inbred individuals could act as more effective reservoirs of infectious agents.

We sampled tissues from 371 stranded California sea lions (*Zalophus californianus*; Fig. 1) that were being rehabilitated at the Marine Mammal Center in Sausalito, California. Each individual was assessed for age category (based on length, pelage, sagittal-crest size and tooth development) and health status. We identified six classes of health problem^{3–6}: trauma (inflicted by boats, gunshot or net trapping), carcinoma, bacterial infection, helminth infection, intoxication by domoic acid as a result of ingesting prey contaminated by the toxin-producing alga *Pseudonitzschia australis*, and 'nonspecific' (used to classify morbid sea lions with subclinical infections⁶). Animals suffering from trauma were considered as controls that were representative of the normal adult population.

All sea lions were scored for 11 highly polymorphic microsatellite DNA markers^{7–10} that were in Hardy–Weinberg equilibrium. Homozygous individuals were genotyped twice to guard against the possibility of allele drop-out. As a measure of inbreeding, we estimated the relatedness of an individual's parents by calculating an 'internal relatedness' factor¹¹.

Mean internal-relatedness values varied significantly among the six different disease classes (analysis of variance (ANOVA): $F_{5,370}=8.97$, $P=10^{-8}$), being lowest for trauma and highest for carcinoma (Fig. 2). A generalized linear model did not reveal any significant effect of sex or age class on internal relatedness and disease.

After trauma, the next-lowest internal-relatedness values are found for individuals affected by algal toxin. The difference in mean internal relatedness between trauma and toxin is significant (two-tailed t -test: $t=2.74$, d.f. = 135, $P=0.007$). Removal of these two non-infectious disease classes does not eliminate interclass differences (ANOVA: $F_{3,233}=5.66$, $P=0.0009$).

For stringent comparison between infectious agents, we also removed carcinoma (although not an infectious disease, it is strongly associated with herpesvirus infection in sea lions⁴) and the nonspecific category because the sources of morbidity were not defined. This leaves bacterial infection (mean internal relatedness, 0.11) and helminth infection (mean internal relatedness, 0.20), for which the difference is significant between both groups (two-tailed t -test: $t=3.07$, d.f. = 168, $P=0.002$). Thus, the extent to which inbreeding increases susceptibility to infectious disease varies depending on the class of pathogen involved.

The trend that is evident in Fig. 2 indicates that an individual's susceptibility to a biochemically simple challenge, such as a toxin, may not be markedly affected by inbreeding; however, inbreeding seems to become an important factor in susceptibility to more complex, long-lived gut parasites that interact with many aspects of the host's biology. To test this possibility, we surveyed faecal eggs to determine the number of different types of helminth infection in affected sea lions, and found a significant positive regression of parasite diversity on inbreeding ($r^2=0.11$, $n=72$, $P=0.003$). This indicates that the more inbred animals harbour a wider range of helminth infections.

To investigate whether a sick individual's parental relatedness also affects its response to treatment, we recorded the recovery times of animals with 'nonspecific' morbidity that were not receiving any treatment against a specific disease. We found a significant positive regression of rehabilitation time on inbreeding ($r^2=0.14$, $n=43$, $P=0.01$), indicating that the more inbred individuals take longer to recover.

All of the challenges (apart from trauma) that we studied were influenced by inbreeding, which supports findings that link



Figure 1 Bad breeding? Inbred sea lions' increased susceptibility to disease could pose a risk to their fellows as well as themselves.

K. ACEVEDO-WHITEHOUSE

heterozygosity of the major histocompatibility complex with an effective response to challenge (see ref. 12, for example).

The importance of genetics in demography and conservation is contentious¹³. Our results indicate that the role of stochasticity in the natural mortality of vertebrates may be less important than was previously thought, and that inbreeding could have a significant impact on conservation programmes and the dynamics of wildlife diseases. The most inbred individuals not only cost more to treat and rehabilitate, but they could also act disproportionately as reservoirs of infectious agents when they are subsequently released.

Karina Acevedo-Whitehouse*, Frances Gulland†, Denise Greig†, William Amos*

*Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK
e-mail: w.amos@zoo.cam.ac.uk

†The Marine Mammal Center, Marin Headlands, 1065 Fort Cronkhite, Sausalito, California 94965, USA

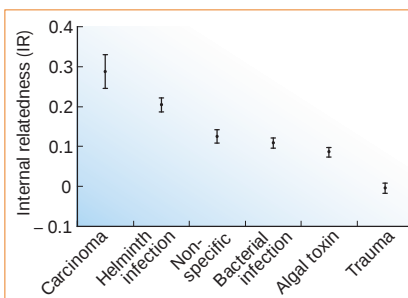


Figure 2 Internal relatedness in sea lions and the incidence of different disease classes. Carcinoma, $n=13$; helminth infection, $n=72$; nonspecific, $n=51$; bacterial infection, $n=98$; algal toxin, $n=101$; trauma (control), $n=36$. Values are means $\pm$ s.e. The mean internal relatedness value of trauma animals (-0.004) is indistinguishable from zero, as would be expected for individuals born to randomly mated parents⁹.

- Keller, F. & Waller, D. M. *Trends Ecol. Evol.* **17**, 230–241 (2002).
- Coltman, D. W., Pilkington, J. G., Smith, J. A. & Pemberton, J. *Evolution* **53**, 1259–1267 (1999).
- Gulland, F. M. D. et al. *Vet. Rec.* **150**, 475–480 (2002).
- Lipscomb, T. P. et al. *Vet. Pathol.* **37**, 609–617 (2000).
- Goldstein, T. et al. *Aquat. Mamm.* **25**, 43–51 (1999).
- Gage, L. J., Gerber, J. A., Smith, D. M. & Morgon, L. E. *J. Zoo Wild. Med.* **24**, 41–47 (1993).
- Coltman, D. W., Bowen, W. D. & Wright, J. M. *Mol. Ecol.* **5**, 161–163 (1996).
- Allen, P., Amos, W., Pomeroy, P. P. & Twiss, S. D. *Mol. Ecol.* **4**, 653–662 (1996).
- Hoelzel, A. R., Campagna, C. & Arnborn, T. *Proc. R. Soc. Lond. B* **268**, 325–332 (2000).
- Buchanan, F. C., Maier, L. D., Thoe, T. D., DeMarch, B. G. & Stewart, R. C. *Mol. Ecol.* **7**, 1083–1085 (1998).
- Amos, W. et al. *Proc. R. Soc. Lond. B* **268**, 2021–2027 (2001).
- Penn, D. J., Damjanovich, K. & Potts, W. K. *Proc. Natl Acad. Sci. USA* **99**, 11260–11264 (2002).
- Frankham, R., Ballou, J. D. & Briscoe, D. A. *Introduction to Conservation Genetics* (Cambridge Univ. Press, Cambridge, UK, 2002).

Competing financial interests: declared none

Psychophysics

Is subliminal learning really passive?

Perceptual learning can occur as a result of exposure to a subliminal stimulus, without the subject having to pay attention and without relevance to the particular task in hand¹ — but is this type of learning purely passive? Here we show that perceptual learning is not passive, but instead results from reinforcement by an independent task^{2,3}. As this learning occurred on a subliminal feature, our results are inconsistent with attentional learning theories^{4,5} in which learning occurs only on stimuli to which attention is directed. Instead, our findings suggest that the successful recognition of a relevant stimulus can trigger an internal reward⁶ and give rise to the learning of irrelevant and even subliminal features that are correlated with the occurrence of the reward^{5,6}.

In our previous study¹, subjects carried out an attentionally demanding letter-identification task⁷ in the fovea while a coherently moving, random-dot display that was below the visibility threshold was presented in the periphery. Repetitive exposure improved subjects' performance in specifically identifying the direction of this subthreshold motion when tested in a subsequent above-threshold test¹. This showed that perceptual learning occurs

unconsciously as a result of frequent exposure to an irrelevant feature.

This raises the question of whether subliminal learning occurs only as a result of passive exposure to a stimulus. To test this idea, we presented four different directions of motion an equal number of times during the exposure stage of the experiment, with a 'designated direction' always being 'paired' with the targets of the letter-identification task and the other directions being randomly associated with the letter-task distractors (Fig. 1a). If perceptual learning were due only to passive exposure, motion direction thresholds should improve equally for all of the presented directions. But if learning occurs only for features to which attention is directed, no improvement should be found for any presented direction.

However, contrary to both situations, the threshold improved for the designated direction paired with the task targets, but not for the other directions, which were paired with distractors (Fig. 1b). Our results indicate learning of an irrelevant and subliminal feature that is positively correlated with target presentation. No evidence of passive learning was found for the other directions of random-dot motion within the exposure period, although we cannot rule out passive learning over longer periods of exposure.

To determine whether the target letters facilitate learning of the designated direction (as in internal reinforcement) or

whether the distractor letters inhibit learning of the other directions of motion (as in negative priming⁸), or both, we carried out a control experiment using new subjects. These new subjects were passively exposed to the stimulus while, as their main task, they were instructed to report the 5% of trials in which they saw the same letter twice in a sequence of black letters. Consistent with the reinforcement learning hypothesis, subjects showed no significant subliminal learning (results not shown). However, the nonsignificant negative trend of distractors in the main study, but not in the control experiment, may indicate that there is a learning-inhibition component.

Although our results are at odds with theories of passive learning and focused attention, they are consistent with classic conditioning results³, in which arbitrary features are learned as a result of pairing with rewarding or noxious stimuli. In our experiment, the motion direction (designated direction) was paired with task targets, the successful recognition of which may generate an internal reward⁶.

We propose that diffuse reinforcement-learning signals are complementary to focused attention^{4,5} in subliminal perceptual learning. In reinforcement learning, a reward signal links the neural response with task performance³, whereas focused attention allows knowledge to bias the neural response⁹ and involves cortical processing. We have shown that presenting a stimulus that is relevant to a task can give rise to an internal reward that works like an external reward in reinforcement learning. This reinforcement signal is probably mediated by neurotransmitters such as acetylcholine, noradrenaline and dopamine from sub-cortical brain areas, which are widely released in a task-specific manner¹⁰ and bring about synaptic plasticity².

Aaron R. Seitz*, Takeo Watanabe†

**Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA*

e-mail: aseitz@hms.harvard.edu

†Department of Psychology, Boston University, Boston, Massachusetts 02215, USA

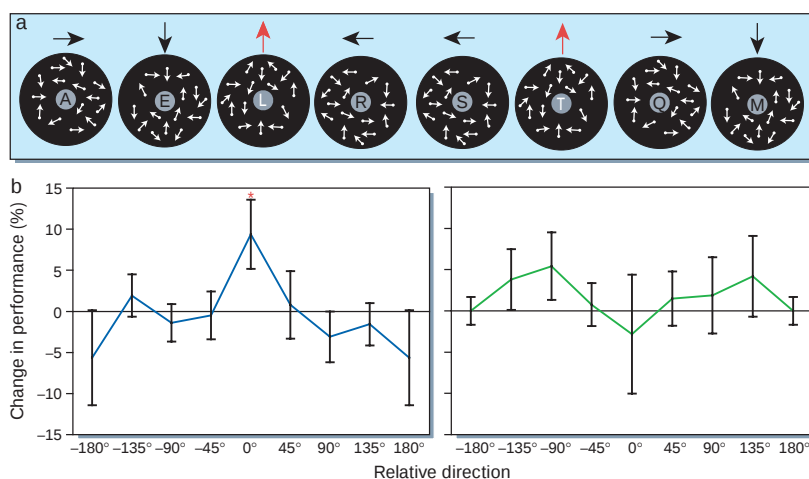


Figure 1 Experimental investigation of perceptual learning, in which an exposure stage was preceded and followed by a test stage¹.

a, Exposure stage, showing the letters used in the main identification task and the direction of motion of background dynamic random-dot displays (DRDs, arrows). Red arrows indicate the 'designated direction' for the motion of the DRD that was presented at the same time as the task targets (two white letters); black arrows indicate the other directions of the DRD, which were paired with the task distractors (six black letters). In the tests¹, subjects reported direction of coherent motion of DRDs of 5% and 10%; their performance in the tests was at chance at 5% coherence and better than chance at 10%. In the exposure stage, subjects witnessed DRDs with 5% coherent motion while carrying out the identification task¹ on dimmed letters to prevent their attention from disengaging. Letters were presented for 375 ms, centred in a 500-ms DRD. **b**, Perceptual-learning results (mean for five subjects), showing the difference in performance in tests before and after exposure for each direction (centred on the designated direction) at 5% and 10% coherent DRDs. At 10% coherence (left), there was a significant peak ($P < 0.01$, analysis of variance) and the identification accuracy for the designated direction (0°) was significantly better than for other directions ($P < 0.05$, one-tailed t -test versus 0°; $P < 0.01$ for designated direction versus other directions) for all subjects. At 5% coherence (right), no significant effect was evident.

1. Watanabe, T., Náñez, J. E. & Sasaki, Y. *Nature* **413**, 844–848 (2001).
 2. Arnsten, A. J. *Neuropharmacology* **11**, 151–162 (1997).
 3. Schultz, W. *Nature Rev. Neurosci.* **1**, 199–207 (2000).
 4. Ahissar, M. & Hochstein, S. *Proc. Natl Acad. Sci. USA* **90**, 5718–5722 (1993).
 5. Karni, A. & Sagi, D. in *Maturation Windows and Adult Cortical Plasticity, SFI Studies in the Sciences of Complexity* (ed. Kovács, B. J. a. I.) 1–15 (Addison-Wesley, Reading, Massachusetts, 1995).
 6. Herzog, M. H. & Fahle, M. *Vision Res.* **39**, 4232–4243 (1999).
 7. Joseph, J. S., Chun, M. M. & Nakayama, K. *Nature* **387**, 805–807 (1997).
 8. Tipper, S. P. & Cranston, M. Q. *J. Exp. Psychol. A* **37**, 591–611 (1985).
 9. Reynolds, J. H., Chelazzi, L. & Desimone, R. *J. Neurosci.* **19**, 1736–1753 (1999).
 10. Dalley, J. W. et al. *J. Neurosci.* **21**, 4908–4914 (2001).
- Competing financial interests:** declared none.

Regulated portals of entry into the cell

Sean D. Conner & Sandra L. Schmid

Department of Cell Biology, The Scripps Research Institute, La Jolla, California 92037, USA

The plasma membrane is the interface between cells and their harsh environment. Uptake of nutrients and all communication among cells and between cells and their environment occurs through this interface. 'Endocytosis' encompasses several diverse mechanisms by which cells internalize macromolecules and particles into transport vesicles derived from the plasma membrane. It controls entry into the cell and has a crucial role in development, the immune response, neurotransmission, intercellular communication, signal transduction, and cellular and organismal homeostasis. As the complexity of molecular interactions governing endocytosis are revealed, it has become increasingly clear that it is tightly coordinated and coupled with overall cell physiology and thus, must be viewed in a broader context than simple vesicular trafficking.

The plasma membrane is a dynamic structure that functions to segregate the chemically distinct intracellular milieu (the cytoplasm) from the extracellular environment by regulating and coordinating the entry and exit of small and large molecules. Essential small molecules, such as amino acids, sugars and ions, can traverse the plasma membrane through the action of integral membrane protein pumps or channels. Macromolecules must be carried into the cell in membrane-bound vesicles derived by the invagination and pinching-off of pieces of the plasma membrane in a process termed endocytosis. Endocytosis occurs by multiple mechanisms that fall into two broad categories, 'phagocytosis' or cell eating (the uptake of large particles) and 'pinocytosis' or cell drinking (the uptake of fluid and solutes). Phagocytosis is typically restricted to specialized mammalian cells, whereas pinocytosis occurs in all cells by at least four basic mechanisms: macropinocytosis, clathrin-mediated endocytosis (CME), caveolae-mediated endocytosis, and clathrin- and caveolae-independent endocytosis (Fig. 1). These mechanistically diverse and highly regulated endocytic pathways function to control such complex physiological processes as hormone-mediated signal transduction, immune surveillance, antigen-presentation, and cellular and organismal homeostasis. The mechanistic complexities that govern endocytosis suggest that great evolutionary effort has been expended to control entry into the cell and thereby to control cellular responses to the environment.

Endocytosis might have been the first vesicular trafficking event to evolve in the prokaryote-to-eukaryote transition. De Duve¹ has suggested that as early cells moved from the concentrated primordial soup to the more dilute environment of nascent oceans, the

existing strategy of secreting digestive enzymes and pumping small catabolic products across the plasma membrane became untenable. Thus, early cells evolved mechanisms to take up extracellular macromolecules (pinocytosis) and secrete digestive enzymes directly into an endocytic vacuole containing the appropriate nutrient pumps (a lysosome progenitor). Subsequently, the acquisition of mitochondria and chloroplasts is believed to have occurred by phagocytosis of eubacteria and/or blue-green algae, respectively. The mechanisms of endocytosis have diversified considerably since its early evolutionary beginnings. In multicellular organisms, distinct endocytic pathways are tightly regulated to control all aspects of intercellular communication. Interestingly, several otherwise mechanistically diverse endocytic pathways are regulated by the large GTPase dynamin, which will be described in some detail. Here we summarize our recent understanding of these pathways and the mechanisms that control endocytic vesicle formation in mammalian cells.

Phagocytosis

Phagocytosis in mammals is conducted primarily by specialized cells, including macrophages, monocytes and neutrophils, that function to clear large pathogens such as bacteria or yeast, or large debris such as the remnants of dead cells, arterial deposits of fat, and so on². It is an active and highly regulated process involving specific cell-surface receptors and signalling cascades mediated by Rho-family GTPases³. For example, Fc receptors on macrophages recognize and are activated by antibodies bound to surface antigens on bacteria. A signalling cascade involving activation of Cdc42 and Rac triggers actin assembly and the formation of cell-surface

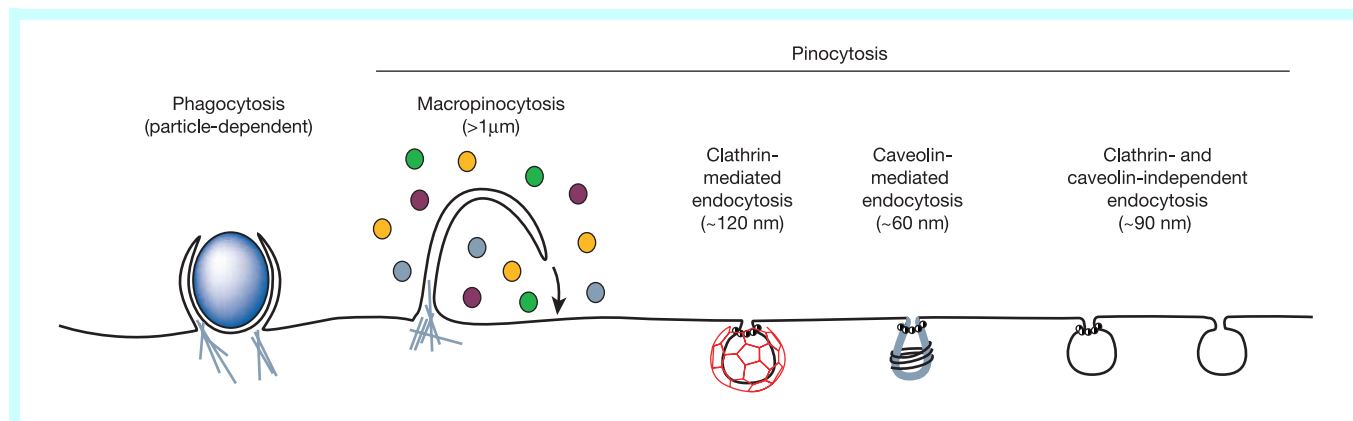


Figure 1 Multiple portals of entry into the mammalian cell. The endocytic pathways differ with regard to the size of the endocytic vesicle, the nature of the cargo (ligands, receptors and lipids) and the mechanism of vesicle formation.

extensions that zipper up around the antibody-coated pathogen to engulf it (Fig. 2a). These signals also activate the cell's inflammatory responses so that once the bacteria are taken up into membrane-bound phagosomes, they are destroyed by the infusion of a barrage of bactericidal weapons, including acids, free oxygen radicals and acid hydrolases. Regurgitated bacterial peptides are then presented on the surface of macrophages to elicit the immune response. Phagocytosis is also crucial for clearing apoptotic cells, both at sites of inflammation and tissue damage and during development. Different receptors on the macrophage surface are involved: for example, the complement receptor or phosphatidylserine receptor that recognize damaged cells. These receptors activate other Rho-family GTPases to trigger membrane extension and phagocytosis, but not the accompanying inflammatory responses⁴. Phagocytosis is inhibited by overexpression of a dominant-negative dynamin mutant⁵, although it is not clear whether this reflects the role of dynamin in vesicle formation or in the regulation of the actin cytoskeleton (see below).

There are multiple modes of phagocytosis, which are determined by the particle to be ingested and the receptor that recognizes that particle. This ensures that phagocytosis operates in conjunction with other aspects of the cell's or organism's physiological response. Some bacteria, like the mycobacteria that cause tuberculosis, use this to their advantage and direct the formation of customized phagocytic vesicles that are ideally suited to house their intracellular

replication⁶. Several excellent reviews describe these diverse and specialized phagocytic mechanisms^{2,7}. Here we will focus on the pinocytic pathways that operate in all mammalian cells.

Multiple pathways for pinocytosis

Pinocytosis, or fluid-phase uptake, can be measured by the intracellular accumulation of tracer molecules (for example, an enzyme or labelled compound) present in the medium. The degree of internalization of fluid-phase markers is directly proportional to their concentration in the medium and the volume encased by the transport vesicles. Greater efficiency of endocytosis is achieved by nonspecific binding of solutes to the cell membrane (adsorptive pinocytosis), but the most efficient uptake occurs when dilute solutes are captured by specific high-affinity receptors (receptor-mediated endocytosis), which are themselves concentrated into specialized endocytic transport vesicles. As for phagocytosis, the cargo molecule and its receptor determine the pinocytic pathway through which they gain entry into the cell.

Macropinocytosis

Macropinocytosis accompanies the membrane ruffling that is induced in many cell types upon stimulation by growth factors or other signals. Like phagocytosis, the signalling cascades that induce macropinocytosis involve Rho-family GTPases, which trigger the actin-driven formation of membrane protrusions. However, unlike phagocytosis, these protrusions do not 'zipper up' along a ligand-coated particle, but instead they collapse onto and fuse with the plasma membrane (Fig. 1) to generate large endocytic vesicles, called macropinosomes, that sample large volumes of the extracellular milieu. Little is known about the nature of this fusion process. For example, how frequently do the contacts between collapsed membrane protrusions and the cell surface result in fusion, and is it efficient? Does fusion occur only at specific contact sites, and how is it regulated? What prevents fusion with neighbouring cells? Although macropinocytosis accompanies seemingly chaotic membrane ruffling, it is likely to be a highly controlled and regulated process.

Macropinocytosis fulfils diverse functions. It can be transiently induced in most cells and might have a role in the downregulation of activated signalling molecules. Platelet-derived growth factor (PDGF)-induced macropinocytosis, which involves activation of the Rho-GTPase Rac and its downstream kinase, p21-activated kinase (PAK), might have a role in directed cell migration⁸. Activation of antigen-presenting dendritic cells triggers extensive and prolonged macropinocytic activity, enabling these cellular sentries to sample large volumes of the extracellular milieu and to fulfil their role in immune surveillance⁹. Finally, some bacteria inject toxins into cells that activate Rho-GTPases, triggering macropinocytosis and ensuring their own uptake into macropinosomes, which are conducive to their replication⁶.

The actin-driven formation of large macropinosomes or phagosomes differs mechanistically from the involution of more selective plasma-membrane domains that give rise to smaller pinocytic vesicles. The remainder of this review focuses on three mechanistically distinct pathways of selective uptake: first, clathrin-mediated endocytosis; second, caveolae-mediated endocytosis; and third, caveolin- and clathrin-independent endocytosis, which probably encompasses more than one pathway.

Caveolae-mediated endocytosis

Caveolae are flask-shaped invaginations of the plasma membrane that were first observed, 50 years ago, on the surface of endothelial cells, where they are extremely abundant. They were proposed to mediate the extensive transcellular shuttling of serum proteins from the bloodstream into tissues across the endothelial cell layer. Caveolae are now known to be present on many cells, and to demarcate cholesterol and sphingolipid-rich microdomains of the

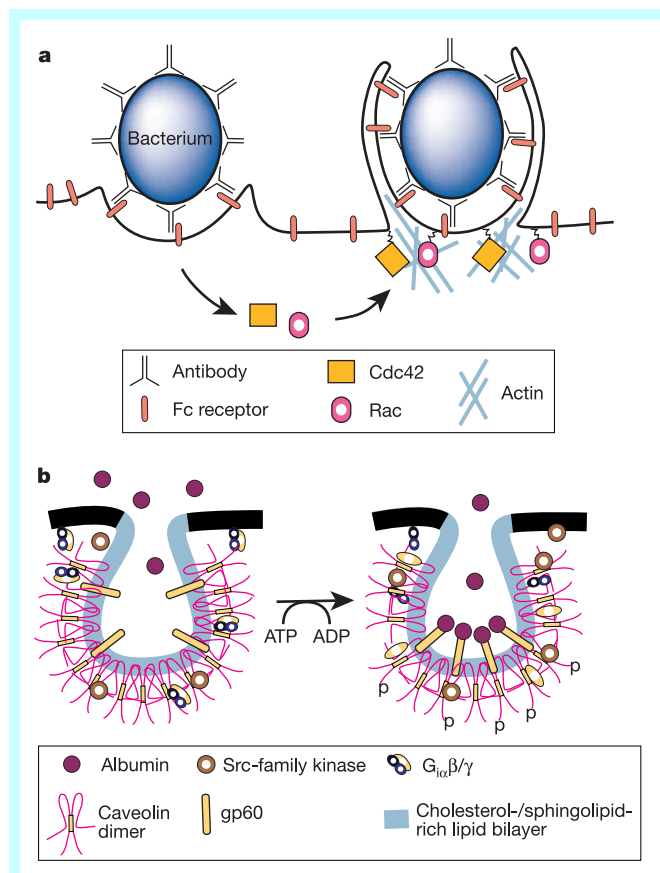


Figure 2 Cargo-stimulated signalling pathways induce uptake by phagocytosis and caveolae. **a**, Fc receptors on the surface of macrophages are activated by immunoglobulin- γ molecules bound to a bacterium. A signalling cascade that involves Rac, Cdc42 and downstream kinases triggers actin rearrangements, protrusion of the membrane around the bacterium, and its engulfment into a phagosome. **b**, Albumin binds to and presumably clusters its receptor, gp60, in caveolae to activate $G_{i\alpha}$ and Src kinases, triggering caveolae endocytosis.

plasma membrane, in which many diverse signalling molecules and membrane transporters are concentrated¹⁰. The shape and structural organization of caveolae are conferred by caveolin, a dimeric protein that binds cholesterol, inserts as a loop into the inner leaflet of the plasma membrane, and self-associates to form a striated caveolin coat on the surface of the membrane invaginations (Fig. 2b). Caveolae have been experimentally disrupted by depletion of plasma-membrane cholesterol, by overexpression of dominant-negative caveolin mutants, and most recently through genetic knockout of the caveolin genes^{11,12}. Caveolin-null mice that lack the main, ubiquitously expressed isoform, caveolin-1, are devoid of morphologically detectable caveolae, establishing its structural importance.

Surprisingly, caveolin-null mice have no overt phenotype; however, several tissue-specific defects have provided important insight into the function of caveolae¹². For example, numerous signalling molecules are associated with caveolae, pointing to a role in the compartmentalization and regulation of specific signalling cascades^{10,12}. In support of this hypothesis, some tissues (for example, lung endothelia) in caveolin-null mice show a hyperproliferative response, and it is clear that caveolae have an important role in negatively regulating endothelial nitric oxide synthase (eNOS) signalling that controls vasodilation. However, given that the mice develop normally and seem not to be susceptible to tumours, a broader role in the negative regulation of signalling cascades seems unlikely. Caveolae are also thought to be involved in intracellular cholesterol trafficking and intracellular cholesterol homeostasis; however, there is no evidence for an imbalance of cellular cholesterol in the caveolin-null mice, so other mechanisms must be in place to ensure cholesterol homeostasis. Caveolae are abundant in adipocytes, and an imbalance in serum lipid levels in caveolin-null mice suggests that the caveolae in these cells might have a broader, organismal role in lipid homeostasis. Further and more detailed analysis of the caveolin knockout mice should continue to provide insight into the function of these organelles.

The role of caveolae and caveolin in endocytosis is more complex than was originally thought. Endothelial cells that lack caveolin are defective in their ability to both bind and take up serum albumin¹², consistent with the proposed role of caveolae in transcellular transport. However, the levels of albumin in both serum and interstitial spaces are normal in caveolin-null mice¹¹. Thus, other pathways for transendothelial transport must exist in the whole organism. Experiments to track caveolae endocytosis in a variety of cultured cells with green fluorescent protein (GFP)-conjugated caveolin show that caveolae are static structures at the plasma membrane^{13,14}. However, their internalization can be triggered through a signalling cascade that results in tyrosine-phosphorylation of caveolae constituents¹⁵. Experimentally, this can be achieved by treating cells with phosphatase inhibitors¹⁶, and more physiologically, by ligands that bind to receptors in caveolae. For example, activation of the serum albumin receptor gp60 triggers caveolae uptake through $G_{i\alpha}$ -dependent activation of the Src tyrosine kinase¹⁷ (Fig. 2b). Opportunistic ligands such as simian virus 40 (SV40) particles can also activate a signalling cascade¹³, which presumably involves the multivalent crosslinking of caveolae-localized surface receptors, to trigger their own uptake. More work is needed to define these signalling cascades, the receptors that activate them and the substrates responsible for caveolae uptake. Interestingly, recent evidence suggests that rather than being required for the endocytosis of caveolae, caveolin might negatively regulate their uptake¹⁸. Although this observation needs further validation, it is consistent with the observed static nature of GFP-caveolin, and with the relatively mild phenotype of caveolin-null mice. Caveolin is tyrosine-phosphorylated by Src *in vivo*¹⁶, but it is not known whether this serves as a signal to release inhibition and trigger internalization. Caveolae-mediated endocytosis can be blocked by either overexpressing dominant-

negative mutants of dynamin or by disrupting actin assembly¹⁵. Thus, both dynamin and dynamic rearrangements of the actin cytoskeleton seem to be required.

In most cells, even after activation, caveolae are only slowly internalized (half-time, $t_{1/2}$, >20 min) and the small vesicles (~50–60 nm in diameter) carry little fluid-phase volume. Thus, it is unlikely that this process contributes significantly to bulk fluid-phase uptake, although the situation is different in endothelia, where caveolae can constitute 10–20% of the cell surface. Rather, caveolae-mediated endocytosis seems to be highly regulated and, as we have seen for phagocytosis and macropinocytosis, driven by the cargo molecules themselves. The molecular basis for this link between cargo molecules, caveolae-localized receptors and triggered endocytosis remains to be elucidated.

Clathrin- and caveolin-independent endocytosis

Caveolae represent just one type of cholesterol-rich microdomain on the plasma membrane. Others, more generally referred to as 'rafts', are small structures, 40–50 nm in diameter, that diffuse freely on the cell surface¹⁹. Their unique lipid composition provides a physical basis for specific sorting of membrane proteins and/or glycolipids based on their transmembrane regions^{19,20}. These small rafts can presumably be captured by, and internalized within any endocytic vesicle. For example, both Shiga toxin and non-aggregated cholera toxins, which bind to raft-associated glycolipids, are internalized by clathrin-coated vesicles (CCVs), together with most bulk membrane markers^{14,21}. By contrast, the interleukin-2 (IL-2) receptor on lymphocytes, which is also associated with lipid microdomains, is internalized in a clathrin- and caveolin-independent manner²². Dominant-negative mutants of Eps15, an adaptor protein complex 2 (AP2) binding partner (see below), that inhibit CME fail to inhibit the internalization of IL-2 receptors, although dominant-negative dynamin mutants do. These alternative endocytic carrier vesicles have been visualized by following GFP-tagged glycosyl phosphatidylinositol (GPI)-anchored proteins²³, but little is known about how they are formed.

Clathrin-independent mechanisms of endocytosis also occur in neurons and neuroendocrine cells, and function in the rapid recovery of membrane proteins after stimulated secretion. Whereas CME is crucial to synaptic function and synaptic vesicle recycling, this slower ($t_{1/2}$ > 1 min) uptake mechanism might predominate only under conditions of high and sustained synaptic activity. Studies on neuroendocrine cells have shown that rapid endocytosis ($t_{1/2}$ < 10 s) after stimulated secretion occurs in a clathrin-independent manner, and involves the neuron-specific isoform of dynamin²⁴. The fact that dynamin mutants inhibit the internalization of caveolae, CCVs and IL-2-containing lipid rafts, but fluid-phase uptake continues, indicates the existence of other, as yet uncharacterized mechanisms for pinocytic uptake. Interestingly, in HeLa cells that express a temperature-sensitive mutant of dynamin, fluid-phase endocytosis is inhibited after a shift to the non-permissive temperature, but is restored within 30–60 min to normal levels by a clathrin- and dynamin-independent mechanism²⁵. The signals leading to upregulation of these alternative endocytic pathways are unknown.

The mechanisms that govern caveolae- and clathrin-independent endocytosis remain poorly understood, as illustrated by the fact that these pathways are described only in negative terms. Nonetheless, it is likely that each of these pathways fulfils unique functions in the cell and varies mechanistically not only in how the vesicles are formed, but in terms of which cargo molecules they transport, to what intracellular destination their cargo is delivered, and how their entry is regulated. It is likely that these different pathways have evolved so that pinocytosis can be coordinated with more complex aspects of cell physiology, such as signal transduction, development and modulation of the cell's responses to and interaction with its environment. Recent identification of cargo molecules specific to

these pathways and of physiological conditions under which they are activated provides important new tools for defining these mechanisms.

Clathrin-mediated endocytosis

The remainder of this review focuses on the better-understood molecular mechanisms governing CME. Some of these (for example, the role of dynamin) will apply directly to other pinocytic pathways. It is also likely that insight into endocytic CCV formation will provide mechanistic paradigms that will guide future analysis of how other pinocytic vesicles form. Finally, elucidation of the molecular links that might coordinate CME with higher-order cellular and organismal functions might provide insight into the physiological basis for the diversity of endocytic pathways.

CME occurs constitutively in all mammalian cells, and carries out the continuous uptake of essential nutrients, such as the cholesterol-laden low-density lipoprotein (LDL) particles that bind to the LDL receptor, and iron-laden transferrin (Tfn) that binds to Tfn receptors^{26,27}. CME was previously referred to as 'receptor-mediated' endocytosis, but it is now clear that this is a misnomer, because most pinocytic pathways involve specific receptor–ligand interactions. CME is crucial for intercellular communication during tissue and organ development^{28,29}, and throughout the life of the organism, as it modulates signal transduction both by controlling the levels of surface signalling receptors, and by mediating the rapid clearance and downregulation of activated signalling receptors. CME is also involved in cell and serum homeostasis by regulating the internalization of membrane pumps that control the transport of small molecules and ions across the plasma membrane, and by recapturing small serum proteins after filtration through the kidney. The clathrin-mediated internalization of voltage-gated calcium channels in neurons helps to control the strength of synaptic transmission, and might have a role in learning and memory³⁰. Finally, CME is required for efficient recycling of synaptic vesicle membrane proteins after neurotransmission³¹.

CME involves the concentration of high-affinity transmembrane receptors and their bound ligands into 'coated pits' on the plasma membrane, which are formed by the assembly of cytosolic coat proteins, the main assembly unit being clathrin. Coated pits invaginate and pinch off to form endocytic vesicles, CCVs, that are encapsulated by a polygonal clathrin coat and carry concentrated receptor–ligand complexes into the cell (Fig. 2). CCVs are very abundant in brain tissue and are relatively easily isolated, allowing identification of the main coat proteins. Clathrin is a three-legged structure, called a triskelion, formed by three clathrin heavy chains, each with a tightly associated clathrin light chain^{27,32} (Fig. 3). Under non-physiological conditions (low salt and high calcium concentrations), clathrin triskelions spontaneously self-assemble into closed polygonal 'cages'. However, clathrin-cage assembly under physiological conditions requires the other main coat constituents, the assembly proteins (APs). Two classes of structurally and functionally distinct APs were identified based on their ability to assemble clathrin^{27,33}: the monomeric assembly protein AP180, and heterotetrameric adaptor protein complexes. There are four structurally related adaptor protein complexes (AP1–4), each mediating vesicle formation at distinct subcellular localizations³⁴; however, only AP2 is involved in endocytic CCV formation. It consists of two large, structurally related subunits called α - and β 2-adaptins, a medium subunit, μ 2, and a small subunit, σ 2 (Fig. 3). AP2 complexes have a barrel-shaped core comprising the amino termini of the adaptin subunits and the two smaller subunits, and two protruding appendages that are reminiscent of 'ears' formed by the carboxy termini of the α - and β 2-adaptins, respectively. Analysis of high-resolution structures of isolated domains of coat proteins, culminating in the recently described structure of the AP2 core domain³⁵, is an active area of research that has provided considerable insight into their assembly and function^{27,36}.

On the basis of its ability to self-assemble into cages and curved lattices, clathrin might be considered as the coat-machinery 'brawn' that drives membrane invagination and vesicle formation. The adaptor protein complexes are the 'brains', as they direct clathrin assembly into curved lattices and couple it to cargo recruitment^{27,33}. The α -adaptin subunit specifies the site of clathrin assembly by targeting AP2 complexes to the plasma membrane, whereas the divergent γ -, δ - and ϵ -adaptin subunits in AP1, 3 and 4 complexes, respectively, target them to other organelles. The β -subunits interact with clathrin and alone are capable of triggering clathrin assembly. The μ 2-subunit binds tyrosine-based internalization motifs on the cytoplasmic domains of endocytic receptors to affect their concentration into coated pits. The σ 2-subunit seems to have a structural role in stabilizing the core domain³⁵. AP180 does not recognize cargo molecules; however, it has the most potent assembly protein activity³⁷ and co-assembles with clathrin into uniform coat structures with a restricted size distribution. AP180 is specifically expressed in neurons, and its functional disruption *in vivo* greatly inhibits synaptic vesicle recycling³⁸. Moreover, the few synaptic vesicles that remain are less uniform in size. Thus, AP180 is thought to be specifically required for efficient clathrin assembly and the recycling of uniformly sized synaptic vesicles in neurons. However, the discoveries of a distantly related AP180 homologue in yeast³⁹ and of a ubiquitously expressed AP180 isoform in mammals, called CALM (for clathrin-assembly lymphoid–myeloid leukaemia)⁴⁰, suggest a broader role for AP180/CALM in CME.

Together, the coat proteins, clathrin, AP2 and AP180, would seem to encode all of the functions necessary to select cargo and form a vesicle. Indeed, cell-free assays have established that the targeted assembly of the coat-protein complexes, COPI and COPII, regulated by the small GTPases Arf and Sar1, seems to be both necessary and sufficient to drive vesicle formation from the Golgi apparatus and the endoplasmic reticulum, respectively^{41,42}. These elegant studies have defined the fundamental mechanisms underlying vesicle formation. By contrast, cell-free assays that have been developed to reconstitute endocytic CCV formation^{43–45} have established that coat proteins are necessary, but not sufficient. Indeed, numerous accessory proteins have been implicated in CME, based

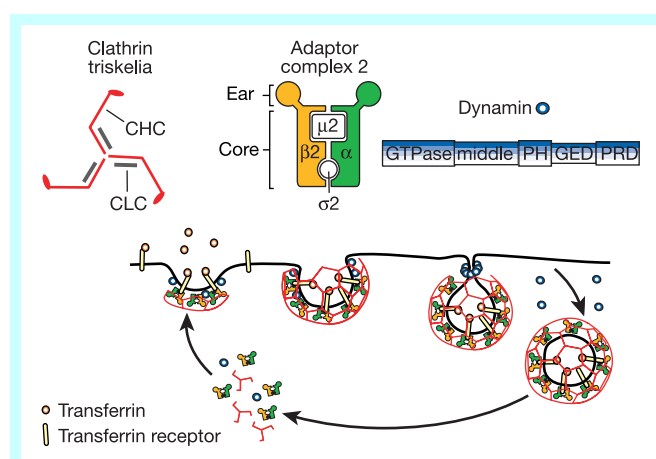


Figure 3 Core components of the machinery driving clathrin-mediated endocytosis. Clathrin triskelions, composed of three clathrin heavy chains (CHC) and three tightly associated light chains (CLC), assemble into a polygonal lattice, which helps to deform the overlying plasma membrane into a coated pit. Heterotetrameric AP2 complexes are targeted to the plasma membrane by the α -adaptin subunits, where they mediate clathrin assembly through the β 2-subunit, and interact directly with sorting motifs on cargo molecules through their μ 2 subunits. Dynamin is a multidomain GTPase that is recruited to the necks of coated pits, where it can assemble into a spiral or 'collar' to mediate or monitor membrane fission and the release of CCVs (see text for details). A subsequent uncoating reaction recycles the coat constituents for reuse.

on their ability to bind AP2, clathrin and/or dynamin. These findings raise several related questions about CCV formation. First, what are the minimum components required to generate a CCV? Second, which of the myriad components of the endocytic apparatus thus far identified have a direct physical role in the formation of the vesicle and which have accessory/regulatory functions? And third, is endocytic CCV formation mechanistically more complex than vesicle formation from intracellular organelles, and if so, why? Answers to these questions will require the identification of the core components of the endocytic apparatus and the function of accessory proteins that impinge on this machinery.

Dynamin regulates endocytic vesicle formation

The GTPase dynamin is required for phagocytosis, caveolae-mediated endocytosis, CME and some clathrin- and caveolae-independent endocytic pathways^{46,47}. Thus, it seems to be a master regulator of membrane trafficking events at the cell surface. The function of dynamin is best characterized in the context of CME.

Dynamin is an atypically large and modular GTPase (Fig. 3) with domains that support phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂) binding (the pleckstrin homology domain, PH), self-assembly (the GTPase effector domain (GED) and the GTPase and middle domains), and interaction with other endocytic components (the proline/arginine-rich domain, PRD). Most intriguingly, dynamin encodes its own GTPase-activating protein (GAP) within the GED. This domain is required for dynamin to self-assemble into helical rings and stacks. Self-assembly can lead to a ~100-fold, GED-mediated stimulation of GTPase activity. At the late stages of CCV formation, dynamin is thought to self-assemble into a 'collar' at the necks of deeply invaginated coated pits (Fig. 3). However, the exact function of dynamin is a matter of some debate⁴⁷, and two models prevail. One suggests that dynamin, unlike any other GTPase-superfamily member, functions as a mechanochemical enzyme to physically drive membrane vesiculation. Two variations on this theme are based on nucleotide-dependent conformational changes undertaken by assembled dynamin molecules. The first, based on the observation that assembled dynamin helices become constricted upon GTP hydrolysis, suggests that dynamin acts as a 'pinchase' or molecular garrotte to constrict and sever invaginated pits at their necks^{48,49}. The second, based on the observation that spirals formed by dynamin in the presence of GDP have a larger helical pitch than those formed in the presence of GTP, suggests that dynamin acts as a 'poppase' or molecular spring to propel the nearly completed vesicle towards the cytosol⁵⁰. The alternative model suggests that dynamin, like other GTPases, functions as a regulatory molecule in endocytic CCV formation to recruit and/or activate downstream effectors while in its GTP-bound form⁵¹.

Regardless of whether dynamin functions as a regulatory GTPase or mechanochemical enzyme, or as some combination of both, there is general agreement that dynamin must undergo GTP-hydrolysis-driven conformational changes for its activity⁵². Indeed overexpression of dominant-negative GTPase domain mutants of dynamin inhibits receptor-mediated endocytosis^{53–55}. The question becomes, does GTP hydrolysis by dynamin function to drive a 'powerstroke', akin to a molecular motor, or does it function to terminate interactions with downstream effectors, akin to a regulatory GTPase? The strongest evidence for the latter is that endocytosis is stimulated by overexpression of GED mutants, defective in self-assembly and consequently in assembly-dependent GAP activity^{51,56}. Further studies are needed to address the role of the assembly-stimulated GTPase activity of dynamin in endocytosis. However, if the assembled dynamin collar is indeed a 'fission machine', then it might be expected that dynamin or some functional counterpart would be fundamentally required for vesicle formation throughout the cell. Thus, it is surprising that mutations in the only dynamin gene in *Drosophila* or *Caenorhabditis elegans*

inhibit endocytic vesicle formation but have not been reported to affect the formation of intracellular transport vesicles. Neither a dynamin-like molecule nor GTP hydrolysis are required for vesicle formation from the ER or Golgi^{41,57}. Finally, dominant-negative mutants of dynamin specifically inhibit endocytosis, and not CCV formation at the trans-Golgi network or delivery of newly synthesized proteins to the plasma membrane in HeLa cells⁵³. Together these data argue against a general requirement for a mechanochemically driven fission apparatus and instead suggest that, like other GTPases, dynamin might have some higher-order function in coordinating endocytic activity with other aspects of cellular function and physiology.

Accessory molecules that regulate CME

A myriad of accessory proteins has been implicated in CME (Fig. 4), initially based on their ability to bind AP2, clathrin and/or dynamin. These have been the subject of recent extensive reviews^{27,58}, and will not be described in detail here. Insight into their specific functions in CCV formation has been gained through *in vivo* analyses after interfering with protein interactions by overexpression of truncated mutants or peptide microinjection, and more recently through *in vitro* analysis of their behaviour on model liposomes^{59–61}. Genetic studies in multicellular organisms have already suggested that some of these proteins (for example, Eps15, endophilin and amphiphysin) might not be essential for CME in all cells. Rather, it is likely that many have regulatory roles that fall into two categories: spatial regulation and/or temporal regulation. Thus, a priority of future studies must be to identify the fundamental mechanisms and the 'core' molecular machinery that drives endocytic CCV formation. This will require new cell-free assays that faithfully reconstitute the hierarchy of molecular interactions leading to CCV formation from biological membranes. Will these fundamentals be different from those established for the formation of other coated vesicles, and if so, why? Once these core components are defined, we can then begin to 'decorate' these coats with the accessory proteins. Here, quantitative kinetic assays will be needed, both *in vivo* and *in vitro*, to define the temporal and spatial coordination and the dynamic interplay between the weakly interacting components whose functions connect distinct cellular processes. This approach will be essential to elucidate the physiological significance of 'accessorizing' endocytic clathrin coats.

Spatial regulation of CME

Scaffolding proteins

Coated pits do not assemble randomly throughout the plasma membrane. Instead, CME is spatially organized at endocytic 'hot-spots' that are, in part, constrained by the actin cytoskeleton⁶². This is especially evident in the neuron, where active zones for synaptic vesicle fusion are encircled by actin-rich regions of high endocytic activity⁶³. A growing list of endocytic accessory proteins with multiple domains for protein–protein and/or protein–lipid interactions can function as scaffolding molecules that connect the endocytic machinery to the actin cytoskeleton, and perhaps more importantly, to other aspects of cell physiology. Among these are amphiphysin, Eps15 and intersectin, which in addition to their interactions with the endocytic machinery, also bind a functionally diverse array of proteins that reflect more complex cellular functions. For example, amphiphysin, which binds AP2, clathrin and dynamin, was originally thought to function in targeting dynamin to coated pits⁶⁴. However, recent amphiphysin knockout studies in mice and *Drosophila* have failed to reveal an effect on CME at the synapse⁶⁵. Instead, amphiphysin seems to function more generally in coordinating membrane microdomains, perhaps through interactions with the actin cytoskeleton. Eps15, which binds AP2 constitutively, has been implicated in regulating coat assembly⁶⁶, although endocytosis is not blocked in Eps15-deficient *C. elegans*.

Interestingly, tyrosine-phosphorylation of Eps15 is essential for EGF receptor internalization, but not for constitutive uptake of Tfn receptors⁶⁷, suggesting that it might have a more subtle role in regulating specific aspects of endocytosis in response to proliferation signals. Studies with model lipid membranes indicate that epsin, the binding partner of Eps15, has a role in vesicle formation by inserting into the bilayer and generating membrane curvature⁶⁸. However, indicative of a broader role in cell physiology, both Eps15 and epsin are transported into the nucleus and might have a role in nucleocytoplasmic transport or in transcriptional regulation^{69,70}. The most striking example of a scaffolding protein that links multiple cellular processes is intersectin⁷¹. Intersectin not only binds components of the endocytic machinery (clathrin, AP2, Eps15, epsin, dynamin and synaptojanin), but also binds to and regulates components of the actin cytoskeleton (neural Wiskott-Aldrich syndrome protein (N-WASP) and Cdc42), components of the Ras/mitogen-activated protein (MAP) kinase signalling pathway (mammalian son-of-sevenless (msos)), and components of the secretory apparatus (secretory carrier membrane proteins (SCAMPs) and the synaptosome-associated protein SNAP-25)^{27,72}. Together, these scaffolding proteins link CME to many other aspects of cell physiology, perhaps providing the means to fine-tune incoming signals and to modulate endocytosis in response to changes in cell physiology.

Lipid interactions

Many components of the endocytic machinery, including AP2, AP180, amphiphysin, epsin, endophilin and dynamin, have domains that interact selectively with the phospholipid PtdIns(4,5)P₂. Several lines of evidence suggest that PtdIns(4,5)P₂ and PtdIns(4,5)P₂ binding are essential for CME. Most notably, many aspects of endocytosis, including coated-pit assembly, CCV formation and their uncoating, are disrupted when PtdIns(4,5)P₂ levels are perturbed by knocking out the phosphatidylinositol-5-phosphatase synaptojanin⁷³. The lipid-binding affinities of most of

these proteins are low, so it is likely that membrane targeting and their lipid-binding functions are enhanced *in vivo* by multivalency and protein-protein interactions.

Lipid interactions might also play a part in generating membrane curvature and perhaps in destabilizing the lipid bilayer to initiate membrane fission. When incubated with PtdIns(4,5)P₂-containing liposomes, amphiphysin, endophilin, epsin and dynamin are each independently able to introduce membrane curvature, converting them into tubular structures^{59,74}. Despite the fact that they interact with PtdIns(4,5)P₂ through structurally distinct domains, both dynamin and epsin insert large portions of themselves into lipid monolayers^{68,75}. Thus, tubulation might reflect a disproportionate increase in the surface area of the outer solvent-exposed leaflet of the lipid bilayer. Asymmetrical insertion into the cytoplasmic leaflet of the plasma membrane could generate curvature during early stages of invagination, and at later stages, could 'strain' or destabilize the bilayer to facilitate membrane fission (Fig. 4). Another mechanism for increasing curvature and destabilizing the lipid bilayer is to directly modify lipids to change their shape and thus their packing in the lipid bilayer. Endophilin has been reported to have lysophosphatidic acid acyl transferase (LPAAT) activity, which adds unsaturated long-chain fatty acids to lysophosphatidic acid, thereby changing it from a 'cone-shaped lipid' to an inverted cone-shaped lipid⁷⁶. Such a transition, if it occurs at the neck of deeply invaginated pits, could assist in membrane fission. However, the LPAAT activity is relatively weak, and only a few lipid molecules would be modified during the time course of vesicle formation. It seems more likely that direct insertion of endophilin into the bilayer would have more dramatic consequences on membrane properties.

The actin cytoskeleton

The actin cytoskeleton is essential for endocytosis in yeast, as demonstrated by the acute disruption of endocytosis after treatment with toxins that disrupt the actin cytoskeleton⁷⁷. By contrast, treatment of mammalian cells with actin-disrupting agents has

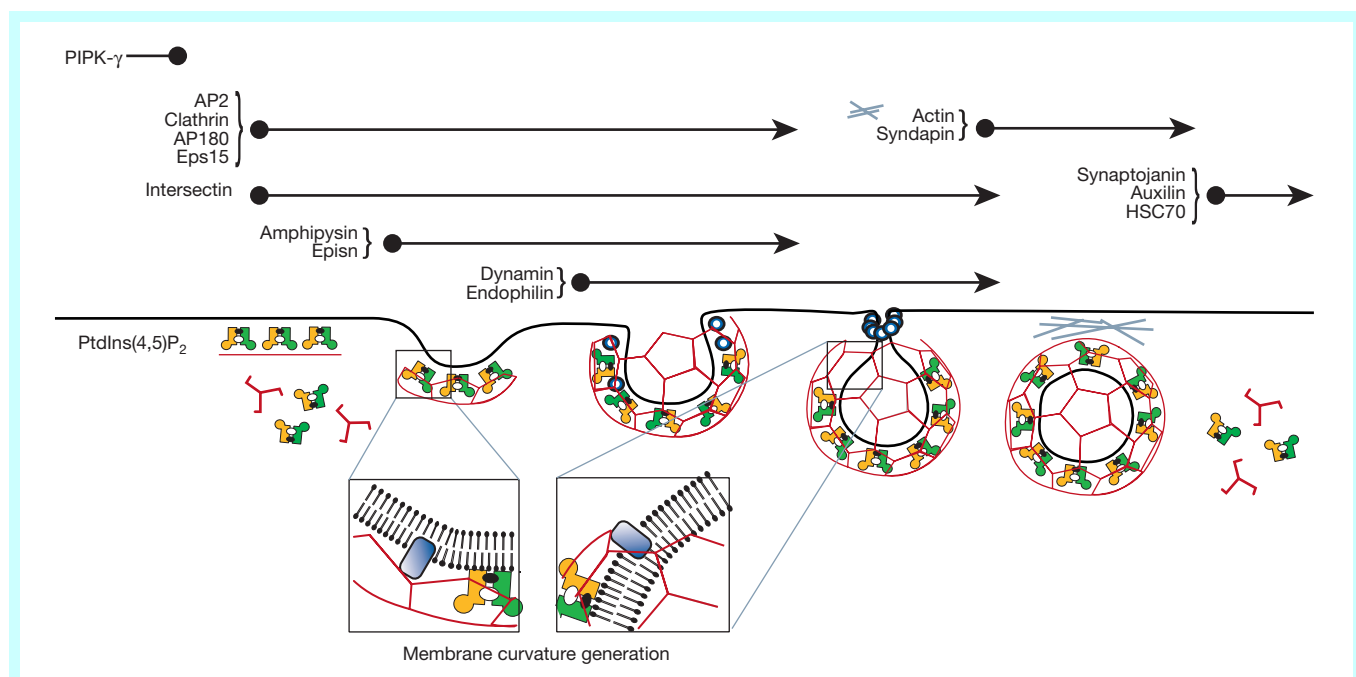


Figure 4 Clathrin-mediated endocytosis is accompanied by the temporally and spatially regulated interactions of multiple factors. The time lines shown are approximate and serve to illustrate the temporal relationships between the dynamic interactions governing CME. Neither the exact hierarchy of these protein interactions nor their exact role in CME is understood. Several accessory proteins involved in

endocytosis, including amphiphysin, endophilin, epsin and dynamin, can induce membrane curvature by asymmetrically inserting a portion of their mass into the outer lipid monolayer (see insets), and might function both early and late in CCV formation. PIPK-γ, phosphatidylinositol-4-phosphate 5-kinase gamma; HSC70, heat shock cognate 70.

only partial or no effect on CCV formation⁷⁸. Nonetheless, many accessory proteins interact directly or indirectly with both the endocytic machinery and the actin cytoskeleton (reviewed in ref. 79). What, then, is the non-essential role of actin in endocytosis in mammalian cells? One possibility, as suggested above, is that actin spatially organizes endocytic hotspots, perhaps serving as an anchor for scaffold assembly. Alternatively, actin assembly might serve to propel nascent endocytic vesicles through the dense cytoskeletal apparatus that lines the cell cortex. Indeed, actin comet tails have been detected on nascent pinocytic vesicles⁸⁰. Recent kinetic data show that actin assembly is spatially and temporally coordinated with the recruitment of dynamin to the necks of coated pits, and with the release of CCVs into the cytosol⁸¹. Moreover, dynamin has been localized to the actin comet tails, and their assembly dynamics are affected by overexpression of dominant-negative dynamin mutants^{82,83}. The role of dynamin and/or other accessory proteins in coordinating these events remains to be established.

Temporal regulation by protein kinases

That CME is spatially and temporally regulated is seen clearly at the synapse, where the role of phosphorylation–dephosphorylation in regulating CCV formation has been most intensively studied. Many components of the endocytic machinery, including dynamin, amphiphysin and AP180, have been identified as ‘dephosphins’, which are phosphorylated in resting neurons and rapidly dephosphorylated by the calcium-dependent phosphatase calcineurin in response to membrane depolarization⁸⁴. Multiple kinases are thought to be responsible for this regulation at the synapse, including protein kinase C, casein kinase II, and others. Several of these co-purify with CCVs that are isolated from brain extracts, and phosphorylate both clathrin and the AP2 complexes. Phosphorylation of the large and medium subunits of AP2 by these ‘endogenous’ CCV kinases can regulate AP2 recruitment to the plasma membrane, AP2 interactions with cargo molecules, and AP2–clathrin co-assembly^{85–87}. The kinase(s) responsible for phosphorylation of α - and β 2-adaptins are unknown, but two distantly related kinases, cyclin-G-associated kinase (GAK, also known as auxilin-2)^{88,89} and adaptor-associated kinase 1 (AAK1)⁹⁰ have been shown to phosphorylate the μ 2-subunit *in vitro*. Phosphorylation by AAK1 of μ 2 on threonine 156, a site that is known to be important for the function of this subunit *in vivo*⁸⁷, markedly increases the affinity of AP2 complexes for tyrosine-based internalization motifs⁹¹. As we learn more about the hierarchy of interactions that coordinate CME, it will be crucial to identify other kinases and phosphatases, and to discover how they regulate coat assembly and vesicle formation.

Cargo selection and regulation of CME

Every pathway of endocytosis that we have discussed seems to be regulated by the nature of the cargo molecule and its receptor. What about the ‘constitutive’ clathrin-mediated pathway? It is now clear that different receptors use different mechanisms for uptake into CCVs. The best-studied endocytic sorting motif, found, for example, on the Tfn receptor, consists of the amino-acid sequence Yx ψ ϕ (where x is any amino acid, ψ a bulky hydrophilic and ϕ a hydrophobic residue). This sequence is recognized by the μ 2-subunit of the AP2 complex³³. The identification of a μ 2-kinase that regulates the affinity of AP2 for these sorting motifs⁹⁰ would imply that this ‘constitutive’ pathway of endocytosis might, in fact, be subject to regulation. Tfn and LDL receptors do not compete for sorting into coated pits⁹², indicating that there are receptor-specific components of the sorting machinery. Indeed, a different sorting motif, FxNPxY, found on the LDL-receptor family, is recognized by Disabled-2 (Dab2), which also binds AP2 and can mediate clathrin assembly^{93,94}. However, a role for Dab2 in regulating LDL-receptor uptake remains to be established. Signalling receptors, like receptor

tyrosine kinases and G-protein-coupled receptors, have also ‘customized’ their uptake mechanisms. The latter use a third adaptor-like molecule, β -arrestin, which also interacts with AP2 and can facilitate clathrin assembly *in vitro*⁹⁵. Both classes of signalling molecules might use ligand-stimulated mono-ubiquitination to post-translationally add new sorting signals and direct their internalization⁹⁶. Endocytosis of signalling receptors, but not of Tfn receptors, can be further modulated by tyrosine-phosphorylation of clathrin⁹⁷, dynamin⁹⁸ and Eps15⁹⁷. Thus, cargo molecules are not passive passengers in these clathrin-coated endocytic vehicles; instead, they might drive their own uptake. This level of regulation, especially for signalling receptors, might have a crucial role in modulating the cell’s response to incoming messages. The numerous molecular connections between signal transduction and endocytosis argue for the physiological importance of these modulatory mechanisms^{28,29,72}.

Future prospects

Evolution has developed diverse mechanisms that are individually and coordinately regulated to control admission into the mammalian cell. Endocytosis is no longer just about ‘eating’ and ‘drinking’. Only by taking a broader ‘systems biology’ approach will we learn how endocytosis affects more complex processes, such as signal transduction, spatial organization, cell migration, polarity, development, and so on—and vice versa. The mechanisms that regulate clathrin-independent endocytosis, and the role of clathrin accessory proteins, must be elucidated. These more subtle regulatory functions are unlikely to be revealed by the overexpression of truncated mutants or other dominant-negative constructs that can wreak havoc because they interfere with the delicate balance of the normal hierarchy of coordinated protein–protein interactions involved. In this regard, new methods of RNA-interference to selectively ‘knock down’ protein expression in mammalian cells will be invaluable. In addition, functional analysis of these proteins in the whole organism, through genetic manipulation, will be essential to determine their specific effects in coordinating endocytic trafficking, modulating intercellular communication, and maintaining tissue and organismal homeostasis. □

doi:10.1038/nature01451.

1. De Duve, C. *Blueprint for a Cell: The Nature and Origin of Life* (Patterson, Burlington, North Carolina, 1991).
2. Aderem, A. & Underhill, D. M. Mechanisms of phagocytosis in macrophages. *Annu. Rev. Immunol.* **17**, 593–623 (1999).
3. Hall, A. & Nobes, C. D. Rho GTPases: Molecular switches that control the organization and dynamics of the actin cytoskeleton. *Phil. Trans. R. Soc. Lond. B* **355**, 965–970 (2000).
4. Fadok, V. A. & Chimini, G. The phagocytosis of apoptotic cells. *Semin. Immunol.* **13**, 365–372 (2001).
5. Gold, E. S. *et al.* Dynamin 2 is required for phagocytosis in macrophages. *J. Exp. Med.* **190**, 1849–1856 (1999).
6. Steele-Mortimer, O., Knodler, L. A. & Finlay, B. B. Poisons, ruffles and rockets: Bacterial pathogens and the host cell cytoskeleton. *Traffic* **1**, 107–118 (2000).
7. Chimini, G. & Chavrier, P. Function of Rho family proteins in actin dynamics during phagocytosis and engulfment. *Nature Cell Biol.* **2**, E191–E196 (2000).
8. Ridley, A. J. Rho proteins: Linking signaling with membrane trafficking. *Traffic* **2**, 303–310 (2001).
9. Mellman, I. & Steinman, R. M. Dendritic cells: Specialized and regulated antigen processing machines. *Cell* **106**, 255–258 (2001).
10. Anderson, R. G. The caveolae membrane system. *Annu. Rev. Biochem.* **67**, 199–225 (1998).
11. Drab, M. *et al.* Loss of caveolae, vascular dysfunction, and pulmonary defects in caveolin-1 gene-disrupted mice. *Science* **293**, 2449–2452 (2001).
12. Razani, B., Woodman, S. E. & Lisanti, M. P. Caveolae: From cell biology to animal physiology. *Pharmacol. Rev.* **54**, 431–467 (2002).
13. Pelkmans, L., Puntener, D. & Helenius, A. Local actin polymerization and dynamin recruitment in SV40-induced internalization of caveolae. *Science* **296**, 535–539 (2002).
14. Thomsen, P., Roepstorff, K., Stahlhut, M. & van Deurs, B. Caveolae are highly immobile plasma membrane microdomains, which are not involved in constitutive endocytic trafficking. *Mol. Biol. Cell* **13**, 238–250 (2002).
15. Pelkmans, L. & Helenius, A. Endocytosis via caveolae. *Traffic* **3**, 311–320 (2002).
16. Parton, R. G., Jørgensen, B. & Simons, K. Regulated internalization of caveolae. *J. Cell Biol.* **127**, 1199–1215 (1994).
17. Minshall, R. D. *et al.* Endothelial cell-surface gp60 activates vesicle formation and trafficking via G γ -coupled Src kinase signaling pathway. *J. Cell Biol.* **150**, 1057–1070 (2000).
18. Le, P. U., Guay, G., Altschuler, Y. & Nabi, I. R. Caveolin-1 is a negative regulator of caveolae-mediated endocytosis to the endoplasmic reticulum. *J. Biol. Chem.* **277**, 3371–3379 (2002).

19. Edidin, M. Shrinking patches and slippery rafts: Scales of domains in the plasma membrane. *Trends Cell Biol.* **11**, 492–496 (2001).
20. Anderson, R. G. & Jacobson, K. A role for lipid shells in targeting proteins to caveolae, rafts, and other lipid domains. *Science* **296**, 1821–1825 (2002).
21. Sandvig, K., Olsnes, S., Brown, J. E., Petersen, O. W. & van Deurs, B. Endocytosis from coated pits of Shiga toxin: A glycolipid-binding protein from *Shigella dysenteriae* 1. *J. Cell Biol.* **108**, 1331–1343 (1989).
22. Lamaze, C. *et al.* Interleukin 2 receptors and detergent-resistant membrane domains define a clathrin-independent endocytic pathway. *Mol. Cell* **7**, 661–671 (2001).
23. Nichols, B. J. & Lippincott-Schwartz, J. Endocytosis without clathrin coats. *Trends Cell Biol.* **11**, 406–412 (2001).
24. Artalejo, C. R., Elhamdani, A. & Palfrey, H. C. Sustained stimulation shifts the mechanism of endocytosis from dynamin-1-dependent rapid endocytosis to clathrin- and dynamin-2-mediated slow endocytosis in chromaffin cells. *Proc. Natl Acad. Sci. USA* **99**, 6358–6363 (2002).
25. Damke, H., Baba, T., van der Bliek, A. M. & Schmid, S. L. Clathrin-independent pinocytosis is induced in cells overexpressing a temperature-sensitive mutant of dynamin. *J. Cell Biol.* **131**, 69–80 (1995).
26. Schmid, S. L. Clathrin-coated vesicle formation and protein sorting: An integrated process. *Annu. Rev. Biochem.* **66**, 511–548 (1997).
27. Brodsky, F. M., Chen, C.-Y., Kneuhl, C., Towler, M. C. & Wakeham, D. E. Biological basket weaving: Formation and function of clathrin-coated vesicles. *Annu. Rev. Cell Dev. Biol.* **17**, 517–568 (2001).
28. Di Fiore, P. P. & De Camilli, P. Endocytosis and signaling: An inseparable partnership. *Cell* **106**, 1–4 (2001).
29. Seto, E. S., Bellen, H. J. & Lloyd, T. E. When cell biology meets development: Endocytic regulation of signaling pathways. *Genes Dev.* **16**, 1314–1336 (2002).
30. Beattie, E. C. *et al.* Regulation of AMPA receptor endocytosis by a signaling mechanism shared with LTD. *Nature Neurosci.* **3**, 1291–1300 (2000).
31. De Camilli, P. & Takei, K. Molecular mechanisms in synaptic vesicle endocytosis and recycling. *Neuron* **16**, 481–486 (1996).
32. Kirchhausen, T. Clathrin. *Annu. Rev. Biochem.* **69**, 699–727 (2000).
33. Kirchhausen, T. Adaptors for clathrin-mediated traffic. *Annu. Rev. Cell Dev. Biol.* **15**, 705–732 (1999).
34. Robinson, M. S. & Bonifacio, J. S. Adaptor-related proteins. *Curr. Opin. Cell Biol.* **13**, 444–453 (2001).
35. Collins, B. M., McCoy, A. J., Kent, H. M., Evans, P. R. & Owen, D. J. Molecular architecture and functional model of the endocytic AP2 complex. *Cell* **109**, 523–535 (2002).
36. Marsh, M. & McMahon, H. T. The structural era of endocytosis. *Science* **285**, 215–220 (1999).
37. Lindner, R. & Ungewickell, E. Clathrin-associated proteins of bovine brain coated vesicles. *J. Biol. Chem.* **267**, 16567–16573 (1992).
38. McMahon, H. T. Endocytosis: An assembly protein for clathrin cages. *Curr. Biol.* **9**, R332–R335 (1999).
39. Wendland, B. & Emr, S. D. Pan1p, yeast eps15, functions as a multivalent adaptor that coordinates protein-protein interactions essential for endocytosis. *J. Cell Biol.* **141**, 71–84 (1998).
40. Tebar, F., Bohlander, S. K. & Sorkin, A. Clathrin assembly lymphoid myeloid leukemia (CALM) protein: Localization in endocytic-coated pits, interactions with clathrin, and the impact of overexpression on clathrin-mediated traffic. *Mol. Biol. Cell* **10**, 2687–2702 (1999).
41. Springer, S., Spang, A. & Schekman, R. A primer on vesicle budding. *Cell* **97**, 145–148 (1999).
42. Rothman, J. E. The machinery and principles of vesicle transport in the cell. *Nature Med.* **8**, 1059–1062 (2002).
43. Lin, H. C., Moore, M. S., Sanan, D. A. & Anderson, R. G. W. Reconstitution of clathrin-coated pit budding from plasma membranes. *J. Cell Biol.* **114**, 881–891 (1991).
44. Smythe, E., Carter, L. L. & Schmid, S. L. Cytosol- and clathrin-dependent stimulation of endocytosis *in vitro* by purified adaptors. *J. Cell Biol.* **119**, 1163–1171 (1992).
45. Gilbert, A., Paccaud, J. P. & Carpentier, J. L. Direct measurement of clathrin-coated vesicle formation using a cell-free assay. *J. Cell Sci.* **110**, 3105–3115 (1997).
46. Hinshaw, J. E. Dynamin and its role in membrane fission. *Annu. Rev. Cell Dev. Biol.* **16**, 483–519 (2000).
47. Sever, S., Damke, H. & Schmid, S. L. Garrotes, springs, ratchets and whips: Putting dynamin models to the test. *Traffic* **1**, 385–392 (2000).
48. Hinshaw, J. E. & Schmid, S. L. Dynamin self assembles into rings suggesting a mechanism for coated vesicle budding. *Nature* **374**, 190–192 (1995).
49. Sweitzer, S. & Hinshaw, J. Dynamin undergoes a GTP-dependent conformational change causing vesiculation. *Cell* **93**, 1021–1029 (1998).
50. Stowell, M. H. B., Marks, B., Wigge, P. & McMahon, H. T. Nucleotide-dependent conformational changes in dynamin: Evidence for a mechanochemical molecular spring. *Nature Cell Biol.* **1**, 27–32 (1999).
51. Sever, S., Muhlberg, A. B. & Schmid, S. L. Impairment of dynamin's GAP domain stimulates receptor-mediated endocytosis. *Nature* **398**, 481–486 (1999).
52. Song, B. D. & Schmid, S. L. A Molecular motor or a regulator? Dynamin's in a class of its own. *Biochemistry* **42**, 1369–1376 (2003).
53. Damke, H., Baba, T., Warnock, D. E. & Schmid, S. L. Induction of mutant dynamin specifically blocks endocytic coated vesicle formation. *J. Cell Biol.* **127**, 915–934 (1994).
54. Damke, H., Binns, D. D., Ueda, H., Schmid, S. L. & Baba, T. Dynamin GTPase domain mutants block endocytic vesicle formation at morphologically distinct stages. *Mol. Biol. Cell* **12**, 2578–2589 (2001).
55. Marks, B. *et al.* GTPase activity of dynamin and resulting conformation change are essential for endocytosis. *Nature* **410**, 231–235 (2001).
56. Sever, S., Damke, H. & Schmid, S. L. Dynamin:GTP controls the formation of constricted coated pits, the rate limiting step in clathrin-mediated endocytosis. *J. Cell Biol.* **150**, 1137–1148 (2000).
57. Rothman, J. E. Mechanisms of intracellular protein transport. *Nature* **372**, 55–63 (1994).
58. Slepnev, V. I. & De Camilli, P. Accessory factors in clathrin-dependent synaptic vesicle endocytosis. *Nature Rev. Neurosci.* **1**, 161–172 (2000).
59. Farsad, K. *et al.* Generation of high curvature membranes mediated by direct endophilin bilayer interactions. *J. Cell Biol.* **155**, 193–200 (2001).
60. Takei, K. *et al.* Generation of coated intermediates of clathrin-mediated endocytosis on protein-free liposomes. *Cell* **94**, 131–141 (1998).
61. Ford, M. G. *et al.* Simultaneous binding of PtdIns(4,5)P₂ and clathrin by AP180 in the nucleation of clathrin lattices on membranes. *Science* **291**, 1051–1055 (2001).
62. Gaidarov, I., Santini, F., Warren, R. A. & Keen, J. H. Spatial control of coated pit dynamics in living cells. *Nature Cell Biol.* **1**, 1–7 (1999).
63. Roos, J. & Kelly, R. B. The endocytic machinery in nerve terminals surrounds sites of exocytosis. *Curr. Biol.* **9**, 1411–1414 (1999).
64. Wigge, P. & McMahon, H. T. The amphiphysin family of proteins and their role in endocytosis at the synapse. *Trends Neurosci.* **21**, 339–344 (1998).
65. Zhang, B. & Zehhof, A. C. Amphiphysins: Raising the BAR for synaptic vesicle recycling and membrane dynamics. *Traffic* **3**, 452–460 (2002).
66. Benmerah, A., Bayrou, M., Cerf-Bensussan, N. & Dautry-Varsat, A. Inhibition of clathrin-coated pit assembly by an Eps15 mutant. *J. Cell Sci.* **112**, 1303–1311 (1999).
67. Confalonieri, S., Salcini, A. E., Puri, C., Tacchetti, C. & Di Fiore, P. P. Tyrosine phosphorylation of Eps15 is required for ligand-regulated, but not constitutive, endocytosis. *J. Cell Biol.* **150**, 905–912 (2000).
68. Ford, M. G. *et al.* Curvature of clathrin-coated pits driven by epsin. *Nature* **419**, 361–366 (2002).
69. Hyman, J., Chen, H., Di Fiore, P. P., De Camilli, P. & Brunger, A. T. Epsin 1 undergoes nucleocytosolic shuttling and its eps15 interactor NH₂-terminal homology (ENTH) domain, structurally similar to Armadillo and HEAT repeats, interacts with the transcription factor promyelocytic leukemia Zn²⁺ finger protein (PLZF). *J. Cell Biol.* **149**, 537–546 (2000).
70. Doria, M. *et al.* The eps15 homology (EH) domain-based interaction between eps15 and hrb connects the molecular machinery of endocytosis to that of nucleocytosolic transport. *J. Cell Biol.* **147**, 1379–1384 (1999).
71. Yamabhai, M. *et al.* Intersectin, a novel adaptor protein with two Eps15 homology and five Src homology 3 domains. *J. Biol. Chem.* **273**, 31401–31407 (1998).
72. McPherson, P. S., Kay, B. K. & Hussain, N. K. Signaling on the endocytic pathway. *Traffic* **2**, 375–384 (2001).
73. Cremona, O. & De Camilli, P. Phosphoinositides in membrane traffic at the synapse. *J. Cell Sci.* **114**, 1041–1052 (2001).
74. Takei, K., Slepnev, V. I., Haucke, V. & De Camilli, P. Functional partnership between amphiphysin and dynamin in clathrin-mediated endocytosis. *Nature Cell Biol.* **1**, 33–39 (1999).
75. Burger, K. N., Demel, R. A., Schmid, S. L. & de Kruijff, B. Dynamin is membrane-active: Lipid insertion is induced by phosphoinositides and phosphatidic acid. *Biochemistry* **39**, 12485–12493 (2000).
76. Schmidt, A. *et al.* Endophilin I mediates synaptic vesicle formation by transfer of arachidonate to lysophosphatidic acid. *Nature* **401**, 133–141 (1999).
77. Ayscough, K. R. Endocytosis and the development of cell polarity in yeast require a dynamic F-actin cytoskeleton. *Curr. Biol.* **10**, 1587–1590 (2000).
78. Fujimoto, J. M., Roth, R., Heuser, J. E. & Schmid, S. L. Actin assembly plays a variable, but not obligatory role in receptor-mediated endocytosis in mammalian cells. *Traffic* **1**, 161–171 (2000).
79. Qualmann, B., Kessels, M. M. & Kelly, R. B. Molecular links between endocytosis and the actin cytoskeleton. *J. Cell Biol.* **150**, F111–F116 (2000).
80. Merrifield, C. J. *et al.* Endocytic vesicles move at the tips of actin tails in cultured mast cells. *Nature Cell Biol.* **1**, 72–74 (1999).
81. Merrifield, C., Feldman, M. E., Wan, L. & Almers, W. Imaging actin and dynamin recruitment during invagination of single clathrin-coated pits. *Nature Cell Biol.* **4**, 691–698 (2002).
82. Lee, E. & De Camilli, P. Dynamin at actin tails. *Proc. Natl Acad. Sci. USA* **99**, 161–166 (2002).
83. Orth, J. D., Krueger, E. W., Cao, H. & McNiven, M. A. The large GTPase dynamin regulates actin comet formation and movement in living cells. *Proc. Natl Acad. Sci. USA* **99**, 167–172 (2002).
84. Cousin, M. A. & Robinson, P. J. The dephosphorylation by calcineurin triggers synaptic vesicle endocytosis. *Trends Neurosci.* **24**, 659–665 (2001).
85. Wilde, A. & Brodsky, F. M. *In vivo* phosphorylation of adaptors regulates their interaction with clathrin. *J. Cell Biol.* **135**, 635–646 (1996).
86. Fingerhut, A., von Figura, K. & Honing, S. Binding of AP2 to sorting signals is modulated by AP2 phosphorylation. *J. Biol. Chem.* **276**, 5476–5482 (2001).
87. Olusanya, O., Andrews, P. D., Swedlow, J. R. & Smythe, E. Phosphorylation of threonine 156 of the μ 2 subunit of the AP2 complex is essential for endocytosis *in vitro* and *in vivo*. *Curr. Biol.* **11**, 896–900 (2001).
88. Greener, T., Zhao, X., Nojima, H., Eisenberg, E. & Greene, L. E. Role of cyclin G-associated kinase in uncoating clathrin-coated vesicles from non-neuronal cells. *J. Biol. Chem.* **275**, 1365–1370 (2000).
89. Korolchuk, V. I. & Banting, G. CK2 and GAK/auxilin2 are major protein kinases in clathrin-coated vesicles. *Traffic* **3**, 428–439 (2002).
90. Conner, S. D. & Schmid, S. L. Identification of an adaptor-associated kinase, AAK1, as a regulator of clathrin-mediated endocytosis. *J. Cell Biol.* **156**, 921–929 (2002).
91. Ricotta, D., Conner, S. D., Schmid, S. L., von Figura, K. & Honing, S. Phosphorylation of the AP2 μ subunit by AAK1 mediates high affinity binding to membrane protein sorting signals. *J. Cell Biol.* **156**, 791–795 (2002).
92. Warren, R. A., Green, F. A., Stenberg, P. E. & Enns, C. A. Distinct saturable pathways for the endocytosis of different tyrosine motifs. *J. Biol. Chem.* **273**, 17056–17063 (1998).
93. Morris, S. M. & Cooper, J. A. Disabled-2 colocalizes with the LDLR in clathrin-coated pits and interacts with AP-2. *Traffic* **2**, 111–123 (2001).
94. Mishra, S. K. *et al.* Disabled-2 exhibits the properties of a cargo-selective endocytic clathrin adaptor. *EMBO J.* **21**, 4915–4926 (2002).
95. Miller, W. E. & Lefkowitz, R. J. Expanding roles for β -arrestins as scaffolds and adaptors in GPCR signaling and trafficking. *Curr. Opin. Cell Biol.* **13**, 139–145 (2001).
96. Hicke, L. A new ticket for entry into budding vesicles—ubiquitin. *Cell* **106**, 527–530 (2001).
97. Wilde, A. *et al.* EGF receptor signaling stimulates SRC kinase phosphorylation of clathrin, influencing clathrin redistribution and EGF uptake. *Cell* **96**, 677–687 (1999).
98. Ahn, S. *et al.* Src-dependent tyrosine phosphorylation regulates dynamin self-assembly and ligand-induced endocytosis of the epidermal growth factor receptor. *J. Biol. Chem.* **277**, 26642–26651 (2002).

Acknowledgements We thank R. Milligan, M. G. Finn and C. Waterman-Storer for critically reading the manuscript, and acknowledge the many researchers whose important contributions to the primary literature can be found in the reviews that were cited owing to space limitations.

Correspondence and requests for materials should be addressed to S.L.S. (e-mail: slschmid@scripps.edu).

Formation of recent martian gullies through melting of extensive water-rich snow deposits

Philip R. Christensen

Department of Geological Sciences, Campus Box 876305, Arizona State University, Tempe, Arizona 85287-6305, USA

The observation of gullies on Mars indicates the presence of liquid water near the surface in recent times^{1,2}, which is difficult to reconcile with the current cold climate. Gullies have been proposed to form through surface runoff from subsurface aquifers^{1,3} or through melting of near-surface ice under warmer conditions⁴. But these gullies are observed to occur preferentially in cold mid-latitudes², where the presence of liquid water is less likely, and on isolated surfaces where groundwater seepage would not be expected, making both potential explanations unsatisfactory. Here I show that gullies can form by the melting of water-rich snow that has been transported from the poles to mid-latitudes during periods of high obliquity within the past 10⁵ to 10⁶ years (refs 5, 6). Melting within this snow⁷ can generate sufficient water to erode gullies in about 5,000 years. My proposed model for gully formation is consistent with the age and location of the gullies, and it explains the occurrence of liquid water in the cold mid-latitudes as well as on isolated surfaces. Remnants of the snowpacks are still present on mid-latitude, pole-facing slopes, and the recent or current occurrence of liquid water within them provides a potential abode for life.

Small young gullies commonly occur in clusters on slopes between 30° and 70° latitude in both hemispheres of Mars². They consist of alcoves several hundred metres wide, and channels up to several kilometres long and several tens of metres deep^{1,2}. Gullies typically originate within several hundred metres of the slope crest, and can occur on crater walls that are raised above the surrounding terrain or near the summit of isolated knobs. Gullies most probably result from flow of liquid water^{1,2}, although alternative fluids—such as concentrated brines^{8,9}, or gaseous or liquid CO₂ (ref. 10)—have been considered.

It has recently been suggested⁴ that periods of high obliquity produce sufficient warming on pole-facing slopes to melt near-surface ice deposits and erode gullies by water-rich debris flows. However, this model does not fully address the fact that as the surface and subsurface temperatures increase, the upper soil layer will become desiccated before significant liquid water can be produced.

Here I propose an alternative model in which melting of an overlying snowpack provides the source of water to erode gullies. The main features of this model are as follows. (1) Water is transported from the poles to mid-latitudes during periods of high obliquity, forming a water-rich snow layer^{5,6,11,12}. (2) Melting occurs at low obliquity as mid-latitude temperatures increase, producing liquid water that is stable beneath an insulating layer of overlying snow. (3) Gullies form on snow-covered slopes through erosion by melt water or as a result of melt water seeping into the loose slope materials and destabilizing them. (4) Gullies incised into the substrate are observed where the snow layer has been completely removed. (5) Patches of snow remain today where they are protected against sublimation by a layer of desiccated dust/sediment¹³. (6) Melting could be occurring at present in favourable locations in these snowpacks.

Evidence for this model is provided by unusual deposits commonly found in the mid-latitudes that may be remnants of an ice-rich mantle^{2,13,14} (Figs 1 and 2). These deposits preferentially occur

on cold, pole-facing slopes, can have features suggestive of flow (Fig. 2), are ~1–10 m thick, have a distinct, rounded edge marking the upslope boundary (Fig. 2), and can occur in hollows at several heights on the same knob (Fig. 1). These characteristics suggest a volatile-rich layer that was once more extensive but which has been removed from all but the cold, pole-facing slopes^{2,14}. The position of the upslope boundary is significant to this model, and probably represents the highest point on the pole-facing slope where ice-rich material is stable under solar illumination.

Figure 3 illustrates the association of gullies with this volatile-rich mantling unit. Patches of smooth mantle remain on the pole-facing (north) interior wall, but gullies incised into the western crater wall are observed where the mantle has been removed. Gullies also occur in the mantle at several locations. The sequence of gully evolution is present in this crater, beginning with a shallow depression in the mantle (northwest wall), progressing to gullies incised into the mantle (north wall), and concluding with gullies incised into the substrate and the snow layer completely gone on the western wall. Figure 4 shows another example of the association of gullies with the mantling unit. Gullies are present where the overlying mantling unit has been removed, but are absent and presumably buried where the mantle remains. Significantly, the uppermost position of the mantle matches the height at which gullies begin, consistent with the mantling unit providing the source of water to carve these gullies.

Several authors have considered the conditions under which snow will accumulate and melt in non-polar regions of Mars^{5,7,11,12,15–17}. Ice removed from the polar regions is transported

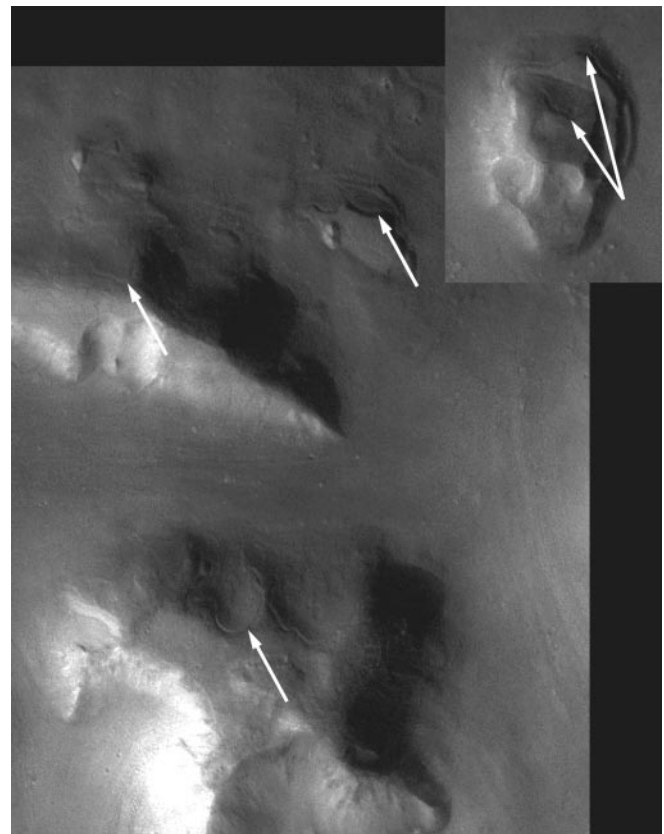


Figure 1 Examples of mid-latitude, pole-facing mantling materials and their curvilinear upper boundaries (arrows). This Mars Odyssey Thermal Emission Imaging System (THEMIS) visible-band image (V01024003) is located at approximately 40.1° N, 350.1° E, and shows the mantling terrains on pole-facing slopes at multiple elevations (arrows). The main portion of the image shown covers an area of about 8.1 × 10.6 km at 18 m per pixel. The inset shows a separate portion of the same image about 5 km to the northeast at the same scale. North is towards the top, and illumination is from the left.

equatorward at rates that are a function of obliquity^{5,6,11,12}, which varies from 15° to >35° on timescales of 10⁵–10⁶ yr (ref. 18). At maximum obliquity, a layer of ice ~0.5–2 cm in thickness could be removed from the summer cap each year and deposited at mid-latitudes^{5,6}. This layer is predominately water, with minor amounts of dust incorporated as ice nucleation centres or accumulated from airfall, similar to the “dirty water ice” found in the north polar residual cap¹⁹. Poleward transport of water to the winter cap will also occur, but at a rate estimated to be 5–50 times slower^{5,6}, resulting in a net accumulation of ~0.1 cm of water ice between 30° and 50° latitude each year. Conditions favourable for water transport last several thousand years during each obliquity cycle, producing a snow layer up to ~10 m thick^{5,6}.

The melting of snow deposits has been modelled by Clow⁷ for differing conditions of snow thickness, atmospheric pressure, snow properties, dust content, and latitude. Melting occurs beneath the surface at temperatures well below freezing, because sunlight is absorbed at depth rather than at the surface, and this absorption is substantially increased by the incorporation of minor amounts of dust⁷. Melting can occur for a wide range of snow properties and atmospheric pressures, and occurs under current conditions in mid-latitudes if dust abundances are greater than ~1,000 parts per million by mass (p.p.m.m.) (ref. 7). These dust-to-ice mixing ratios are readily attained in the present martian atmosphere, and are comparable to the polar ice deposits¹⁹. Meltwater moving

downwards under gravity will encounter lower temperatures and refreeze. However, conduction and latent heat transfer will gradually warm the snow and substrate, allowing liquid water to accumulate and be available for erosion⁷. Subsurface erosion and collapse of the snow mantle will occur, with liquid water potentially reaching and eroding the substrate as the snow layer continues to melt. Liquid water will begin to be generated ~100 d after the spring equinox under current conditions for snow with a dust content of 1,000 p.p.m.m., and will reach a depth of 20 cm approximately 25 d later⁷. Up to 0.33 mm of snowmelt runoff is produced each day for ~50 d each martian year⁷.

Typical alcove size and spacing (~200 × 200 m) provides ~4 × 10⁴ m² of source snowpack for each gully, with melting of 0.25 mm per day producing ~10 m³ of liquid water feeding each gully. For typical gully sizes of 20 m average width, 500 m length, and 10 m depth, a total of 1 × 10⁵ m³ of sediment has been removed from each gully. Gullies occur on steep slopes, and typically cut into what appear to be unconsolidated materials that have accumulated on these slopes. Freeze/thaw action under snow would further the breakdown of the substrate materials. Loose sediment can be transported by runoff with a water-to-sediment ratio of ~10:1 (ref. 20); transport by debris flow requires significantly less water, whereas erosion of bedrock would require more. Assuming loose debris, the total volume of gully material could be removed in 10⁵ days of runoff or ~5,000 elapsed years, and gullies could readily form within the 500,000 yr since the last period of high obliquity. Multiple cycles of snow accumulation and gully erosion probably occur, reinvigorating gullies and expanding their size over several obliquity cycles. There is a strong correlation between the latitudes where ice-rich mantles and gullies occur. However, it is not expected that all snow accumulation sites will have gullies. Gullies would not form where snow accumulation is too thin, the surface too cold, or snow removal too rapid to allow melt to reach the substrate before the snow is removed. Gullies may also remain hidden beneath remnant snow. Although pole-facing, mid-latitude slopes are favoured, gullies could form anywhere that snow accumulates and melts.

Melting of snow can produce a variety of landforms. Water/ice/sediment slurries may develop and catastrophically fail, producing the levees and other features observed on some gullies that are reminiscent of debris flows⁴. Depositional fans and levees may form

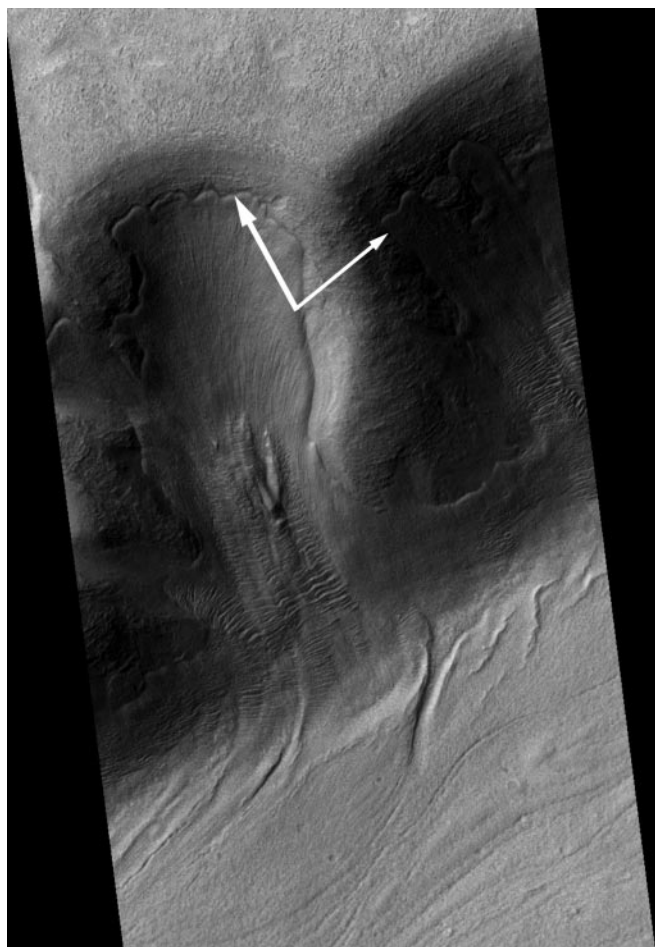


Figure 2 Detail of the appearance and curvilinear upper boundary (arrows) of pole-facing mantling materials located in Dao Valles. The surface morphology of the mantle is suggestive of flow. This Mars Global Surveyor Mars Orbiter Camera (MOC) image (M03-04950²⁶) is centred at 35.7° S, 90.9° E. The portion of the image shown covers an area of about 2.8 × 5 km at 2.8 m per pixel; north is towards the top, the pole is towards the bottom, and illumination is from the left.

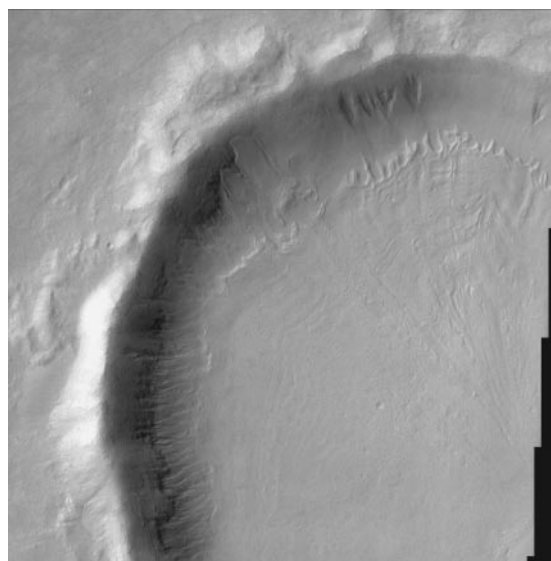


Figure 3 THEMIS image showing gullies emerging from beneath and within a mantling snow unit where this unit is being removed. This portion of a THEMIS image (V01229004; 43.0° S, 214.4° E) covers an area of about 16.2 × 16.2 km at 18 m per pixel; north is towards the top, illumination is from the left.



Figure 4 MOC image of gullies with a remnant of the snow mantle (arrow) proposed to be the source of water that eroded these gullies. This portion of image M09-2875²⁶ covers an area of about 2.8×4.5 km at 2.8 m per pixel near 33.3° S, 92.9° E; north is towards the top, illumination is from the left.

beneath the relatively thin snowpack, or slurries may travel down-slope and deposit on the snow mantle or on the substrate beyond where the snow is stable. Erosion and gully formation can occur within the snow mantle itself, or overlying snow may collapse into buried gullies (Fig. 3).

Snowmelt can explain many of the unusual gully characteristics. It is not necessary to maintain or recharge a near-surface aquifer system; the source of the water comes from an overlying, atmospherically derived condensate, and the recent age reflects recent climate fluctuations capable of producing surficial water-ice deposits^{5–7,11,17}. The overlying snow provides thermal insulation and a barrier against vapour diffusion, allowing liquid water to form and carve the observed gullies. The initiation of gullies within several hundred metres of crater and cliff rims^{1,2} is related to the distance below the rim where snow is stable against sublimation during melting and runoff (Figs 2 and 4). Melting of a draping snow layer accounts for the development of gullies near the summit of isolated knobs where subsurface aquifers are lacking. Finally, gullies occur in clusters where water-ice accumulation and subsequent melting are optimized; water-ice is less likely to accumulate in equatorial zones^{7,13}, and high latitudes ($>70^\circ$) may be too cold to produce melting and gully formation⁷.

In the model proposed here at least some mid-latitude mantles were deposited on the surface with high ice-to-soil ratios. High ($>50\%$ volume) water-ice abundances have been found in the upper few metres at high latitudes^{21–23}, suggesting that this ice also formed at the surface rather than in pores. A pervasive surface texture found from 30° to 50° in both hemispheres has been interpreted to result from ice-cemented soils, 1–10 m thick, that have formed recently and are currently being devolatilized and eroded¹³. Ice is proposed to form through vapour diffusion into pore space^{13,24}. However, the poleward transition from a dissected to continuous surface on this mantle corresponds to a sharp increase in near-surface ice abundance²⁵, suggesting that the mid-latitude portion of this mantle may be the same ice-rich material whose upper few metres have been thoroughly desiccated. In this case these

terrains may have substantially more water than previously suggested¹³.

The model presented here predicts that most gullied surfaces will not be sites of near-surface water reservoirs because the source snow is now gone. In addition, gully formation by snowmelt should not produce salt or mineral deposits, a prediction opposite to what might be expected if the source water was subsurface brines⁹. Finally, the near-freezing temperatures and short duration of gully formation are not likely to cause significant chemical weathering or mineralization of the surface materials.

Snowmelt has important implications for the search for life on Mars and the potential for human exploration. Liquid water is produced within several tens of centimetres of the surface, is stable for extended periods of time, and is regenerated on relatively short (10^5 – 10^6 yr) timescales associated with the variations in orbital parameters. Repeated access to near-surface liquid water could provide a means for life to have survived over extended periods of martian history, and could provide favourable sites for extant life today. The potential for near-surface liquid water on Mars, currently or in the recent past, within centimetres of the surface in remnant snow deposits could influence the rationale for the search for future landing and sample return sites. The rock layers above gullies, previously thought to be the water source, are extremely difficult to access and are unlikely landing sites. However, the pole-facing mantles provide an excellent opportunity to sample and study water-rich units. □

Received 4 September 2002; accepted 14 January 2003; doi:10.1038/nature01436.
Published online 19 February 2003.

- Malin, M. C. & Edgett, K. S. Evidence for recent ground water seepage and surface runoff on Mars. *Science* **288**, 2330–2335 (2000).
- Malin, M. C. & Edgett, K. S. Mars Global Surveyor Mars Orbiter Camera: Interplanetary cruise through primary mission. *J. Geophys. Res.* **106**, 23429–23570 (2001).
- Mellon, M. T. & Phillips, R. J. Recent gullies on Mars and the source of liquid water. *J. Geophys. Res.* **106**, 23165–23180 (2001).
- Costard, F., Forget, F., Mangold, N. & Peulvast, J. P. Formation of recent martian debris flows by melting of near-surface ground ice at high obliquity. *Science* **295**, 110–113 (2002).
- Jakosky, B. M. & Carr, M. A. Possible precipitation of ice at low latitudes of Mars during periods of high obliquity. *Nature* **315**, 559–561 (1985).
- Jakosky, B. M., Henderson, B. G. & Mellon, M. T. Chaotic obliquity and the nature of the Martian climate. *J. Geophys. Res.* **100**, 1579–1584 (1995).
- Clow, G. D. Generation of liquid water on Mars through the melting of a dusty snowpack. *Icarus* **72**, 95–127 (1987).
- Knauth, L. P., Klonowski, S. & Burt, D. Ideas about the surface runoff features on Mars. *Science* **290**, 711–712 (2000).
- Knauth, L. P., Burt, D. M. & Tyburczy, J. A. in *Conference on the Geophysical Detection of Subsurface Water on Mars* 60–61 (Lunar and Planetary Institute Contribution No. 1095, Houston, Texas, 2001).
- Musselwhite, D. S., Swindle, T. D. & Lunine, J. I. Liquid CO_2 breakout and the formation of recent small channels on Mars. *Geophys. Res. Lett.* **28**, 1283–1285 (2001).
- Toon, O. B., Pollack, J. B., Ward, W., Burns, J. A. & Bilski, K. The astronomical theory of climate change on Mars. *Icarus* **44**, 552–607 (1980).
- Haberle, R. M. *et al.* On the possibility of liquid water on present-day Mars. *J. Geophys. Res.* **106**, 23317–23326 (2001).
- Mustard, J. F., Cooper, C. D. & Rifkin, M. K. Evidence for recent climate change on Mars from the identification of youthful near-surface ground ice. *Nature* **412**, 411–414 (2001).
- Carr, M. H. Mars Global Surveyor observations of fretted terrain. *J. Geophys. Res.* **106**, 23571–23595 (2001).
- Gulick, V. C., Tyler, D., McKay, C. P. & Haberle, R. M. Episodic ocean-induced CO_2 greenhouse on Mars: Implications for fluvial valley formation. *Icarus* **130**, 68–86 (1997).
- Lee, P., Cockell, C. S., Marinova, M. M., McKay, C. P. & Rice, J. W. Snow and ice melt flow features in Devon Island, Nunavut, Arctic Canada as possible analogs for recent slope flow features on Mars. (abstract) *Lunar Planet. Sci.* [CD-ROM] 1809 (2001).
- Hecht, M. H. Metastability of liquid water on Mars. *Icarus* **156**, 373–386 (2002).
- Ward, W. R. Present obliquity oscillations of Mars: Fourth-order accuracy in orbital E and I. *J. Geophys. Res.* **84**, 237–241 (1979).
- Kieffer, H. H., Chase, S. C., Jr, Martin, T. Z., Miner, E. D. & Palluconi, F. D. Martian North Pole summer temperatures: Dirty water ice. *Science* **194**, 1341–1344 (1976).
- Howard, A. D. & C. F. McLane, I. Erosion of cohesionless sediment by groundwater seepage. *Water Resour. Res.* **24**, 1659–1674 (1988).
- Feldman, W. C. *et al.* Global distribution of neutrons from Mars: Results from Mars Odyssey. *Science* **297**, 75–78 (2002).
- Mitrofanov, I. *et al.* Maps of subsurface hydrogen from the high energy neutron detector, Mars Odyssey. *Science* **297**, 78–81 (2002).
- Boynton, W. V. *et al.* Distribution of hydrogen in the near surface of Mars: Evidence for subsurface ice deposits. *Science* **297**, 81–85 (2002).
- Mellon, M. T. & Jakosky, B. M. The distribution and behavior of martian ground ice during past and present epochs. *J. Geophys. Res.* **100**, 11781–11799 (1995).

25. Tokar, R. L. *et al.* Ice concentration and distribution near the South Pole of Mars: Synthesis of Odyssey and Global Surveyor analyses. *J. Geophys. Res.* **29**, doi:1029/2002GL015691 (2002).
26. Malin, M. C. *et al.* Image M17-00423, Malin Space Science Systems Mars Orbiter Camera Image Gallery at (http://www.msss.com/moc_gallery) (2002).

Acknowledgements I thank the THEMIS instrument development team at Arizona State University and Raytheon Santa Barbara Remote Sensing, and the spacecraft development and operations teams at the Jet Propulsion Laboratory and Lockheed Martin Astronautics, for the development and performance of the THEMIS instrument and the Odyssey spacecraft. I also thank J. Rice, S. Ruff, H. Kieffer, M. Malin and B. Jakosky for discussions and contributions, and M. Carr and J. Mustard for comments that improved the original manuscript. This work was supported through the NASA Mars Odyssey Project and the NASA Mars Data Analysis Program.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to the author (e-mail: phil.christensen@asu.edu).

Three-dimensional imaging of atomic four-body processes

M. Schulz[†]*, R. Moshhammer*, D. Fischer*, H. Kollmus*, D. H. Madison[†], S. Jones[†] & J. Ullrich*

* Max-Planck Institut für Kernphysik, Saupfercheckweg 1, D 69117 Heidelberg, Germany

† University of Missouri-Rolla, Physics Department and Laboratory for Atomic, Molecular, and Optical Research, Rolla, Missouri 65409, USA

To understand the physical processes that occur in nature we need to obtain a solid concept about the ‘fundamental’ forces acting between pairs of elementary particles. It is also necessary to describe the temporal and spatial evolution of many mutually interacting particles under the influence of these forces. This latter step, known as the few-body problem, remains an important unsolved problem in physics. Experiments involving atomic collisions represent a useful testing ground for studying the few-body problem. For the single ionization of a helium atom by charged particle impact, kinematically complete experiments have been performed^{1–6} since 1969 (ref. 7). The theoretical analysis of such experiments was thought to yield a complete picture of the basic features of the collision process, at least for large collision energies^{8–14}. These conclusions are, however, almost exclusively based on studies of restricted electron-emission geometries^{1–3}. Here, we report three-dimensional images of the complete electron emission pattern for the single ionization of helium by the impact of C^{6+} ions of energy 100 MeV per a.m.u. (a four-body system) and observe features that have not been predicted by any published theoretical model. We propose a higher-order ionization mechanism, involving the interaction between the projectile and the target nucleus, to explain these features.

The experimental set-up¹⁵ is shown schematically in Fig. 1. C^{6+} projectiles of 100 MeV per a.m.u. collided with cold ($<1K$) He gas target atoms from a supersonic jet. The recoiling target ions and the ionized electrons were extracted by a weak electric field and detected by two-dimensional position-sensitive channel plate detectors. A uniform magnetic field of 12 G ($1G = 10^{-4}T$) parallel to the electric field forced the electrons into cyclotron motion (green spiral line). It thereby confined their motion in the plane perpendicular to the extraction field (transverse plane), so that all electrons with a transverse momentum of less than 2 atomic units (a.u.) hit the detector. Both particles were measured in coincidence with the projectiles, which did not change charge state. The momentum components of the recoil ion and the electron in the direction of the extraction field (longitudinal components) were determined from

the time of flight of each particle from the collision region to the respective detector. The transverse momenta were obtained from the position information provided by the detectors for a fixed time of flight.

In Fig. 2a the measured three-dimensional fully differential cross-sections (FDCS) are shown as a function of the azimuthal (φ_e) and polar (θ_e) electron angles. Although the data are only shown for a momentum transfer (defined as the difference between the initial and scattered projectile momentum) of $q = 0.75$ a.u. and an ionized electron energy $E_e = 6.5$ eV, all other final electron energies $E_e < 50$ eV and a large fraction of all possible momentum transfers are recorded simultaneously but not discussed in this letter. The initial projectile direction is along the z axis and $\mathbf{q}$ is pointing nearly in the x direction. φ_e is measured in the x – y plane relative to the x axis, that is $\varphi_e = 0^\circ$ and 180° correspond to the scattering plane (the plane spanned by the initial projectile momentum and the momentum transfer vectors). θ_e is the angle of the electron momentum vector relative to the z axis.

The characteristic double-lobe structure with peaks at ($\theta_e = 90^\circ$, $\varphi_e = 0^\circ$) and ($\theta_e = 90^\circ$, $\varphi_e = 180^\circ$) well-known from high-energy electron-impact studies in the scattering plane^{1–3} is clearly observable. The ($\theta_e = 90^\circ$, $\varphi_e = 0^\circ$) maximum (called the binary peak), which occurs in the direction of $\mathbf{q}$, is relatively easy to understand. It is due to collisions which are dominated by a binary interaction between the projectile and the electron, that is, the recoil ion remains essentially passive. Momentum conservation then demands that the electron be emitted approximately in the direction of $\mathbf{q}$. The ($\theta_e = 90^\circ$, $\varphi_e = 180^\circ$) lobe (called the recoil peak), has been interpreted as a double scattering process¹⁶: the electron, initially emitted approximately in the direction of $\mathbf{q}$, on its path out of the atom elastically backscatters from the recoil ion, which picks up most of the momentum transferred from the projectile.

We expect the ionization cross-sections at the small perturbations Z_p/v_p (where Z_p and v_p are the atomic number and the velocity of the projectile) studied in this work to be essentially identical for electron and ion impact. Furthermore, the results of various theoretical approaches converge with each other with decreasing perturbation. For our collision system, it is therefore not critical which model is used for comparison with the data in order to identify potential deviations from our current understanding. We chose a continuum distorted wave approach with Hartree–Fock wavefunctions for the active electron (CDW–HF), because it is the most powerful model currently available for ion impact ionization.

Figure 2b shows our theoretical FDCS for the same conditions as the experimental data. This CDW–HF calculation is based on refs 10

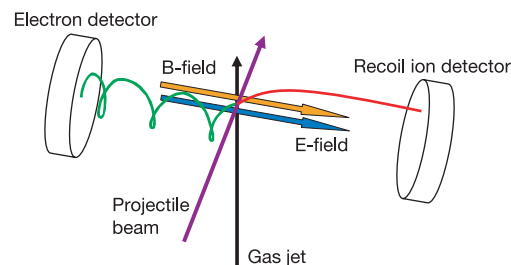


Figure 1 Schematic experimental set-up for three-dimensional imaging of atomic processes. The projectile-ion beam (purple arrow) is crossed with an atomic He beam from a supersonic gas jet (black arrow). The ionized electrons and the recoiling target ions are extracted by a weak electric field (blue arrow) and detected by two-dimensional position-sensitive detectors. A pair of Helmholtz coils generates a uniform magnetic field (B-field, yellow arrow) which forces the electrons into cyclotron trajectories (green spiral line) and guides them onto the detector. Two electronic clocks are used to measure the time of flight of the electrons and the recoil ions from the collision region to the respective detector.

and 11. In our model^{12,13}, the initial state of the collision is represented by a plane wave for the projectile and a Hartree–Fock wavefunction for the He ground state of the electrons. The final state is given by a product of three terms—one for each two-particle subsystem of the final three-particle state. A Coulomb wave is used for the projectile–recoil ion subsystem (that is, a C^{6+} ion in the field of a He^+ ion) and a Coulomb interaction is used for the projectile–electron subsystem. A numerical solution of the Schrödinger equation for an electron in the field of a He^+ ion represented by a static Hartree–Fock potential is employed for the electron–recoil ion subsystem. Although this model represents a state-of-the-art calculation for single ionization by ion impact, significant qualitative differences to the experimental data are apparent. In particular, the calculation is nearly cylindrically symmetric about the x axis (momentum transfer direction) while the data are much broader in the x – y plane than in the x – z plane.

If the ionization process is dominated by a single projectile–target interaction, the symmetry properties of the initial $1s^2\ ^1S$ state lead to cylindrical symmetry about $\mathbf{q}$ in the FDCS. Any deviation from that symmetry is due to higher-order contributions in the interaction between the projectile and the target. The fact that our calculation is nearly cylindrically symmetric shows that in our model, higher-order processes are insignificant. In contrast, the pronounced deviation in the experimental data from the symmetry, indicates that higher-order contributions are important.

To study the discrepancies between the data and the calculation in more detail, we show in Fig. 3 the experimental (solid symbols) and theoretical (solid curves) FDCS for two different cuts through the three-dimensional emission patterns of Fig. 2. The first cut selects

the scattering plane (blue line in Fig. 2) and the second cut the plane perpendicular to $\mathbf{q}$ (red line in Fig. 2). For each selected plane, the FDCS are plotted as a function of the electron emission angle θ_e measured relative to the initial projectile momentum $\mathbf{p}_0$. For the scattering plane, $\theta_e = 90^\circ$ corresponds to the direction of $\mathbf{q}$ and $\theta_e = 270^\circ$ to $-\mathbf{q}$ (note that here θ_e is a plane polar coordinate as opposed to the spherical polar angle used in Fig. 2).

In the scattering plane, the calculation is in good agreement with the data except for some quantitative differences in the recoil peak. It should be noted that both the data and the calculation are absolute, that is, there is no adjustment in the absolute magnitude. Such a good reproduction of the data by theory in the scattering plane is quite typical for single ionization of He by high-energy electron impact even when compared to less sophisticated models such as the first Born approximation¹⁷. In that sense, our results are consistent with high-energy electron-impact data. In the perpendicular plane, in contrast, large discrepancies between experiment and theory are obvious. Here, the theory is essentially isotropic and does not show the peak structures perpendicular to the beam direction, which are still fairly pronounced in the data. Such an isotropic emission pattern for the perpendicular plane is again characteristic of a first-order ionization process. Therefore, the structures observed in the data further support our conclusion above concerning the strong contributions from higher-order processes.

A possible explanation for the discrepancies between our experimental and theoretical results is contained in the perpendicular plane FDCS measured for 2 MeV per a.m.u. $C^{6+} + He$ collisions. These cross-sections are shown in Fig. 4 for $q = 0.7$ a.u. and $E_e = 4$ eV (open symbols) and for $q = 1.5$ a.u. and $E_e = 1$ eV (solid symbols). In the first case, the maxima near 90° and 270° observed for the larger projectile energy are completely absent. In the second case, the electron momentum (0.27 a.u.) is very small compared with q so that here the momentum exchange in the collision occurs predominantly between the projectile and the recoil ion. Under these conditions we observe maxima near 90° and 270° , as we do for the large projectile energy.

Comparison of the data in Fig. 4 provides a strong indication that the peak structures perpendicular to the projectile beam observed in the perpendicular plane are due to a higher-order process involving the projectile–target nucleus interaction. We propose the following two-step mechanism to explain the data; in the first step, the electron is ejected owing to a binary projectile–electron interaction approximately in the direction of the momentum $\mathbf{q}'$ transferred in this initial interaction. The projectile then elastically scatters from the target nucleus. In this second step, an additional momentum $\mathbf{q}''$ is transferred, which may be pointing out of the initial scattering plane ($\mathbf{p}_0, \mathbf{q}'$). The total momentum transfer $\mathbf{q} = \mathbf{q}' + \mathbf{q}''$ measured in the experiment is then rotated about the projectile beam axis

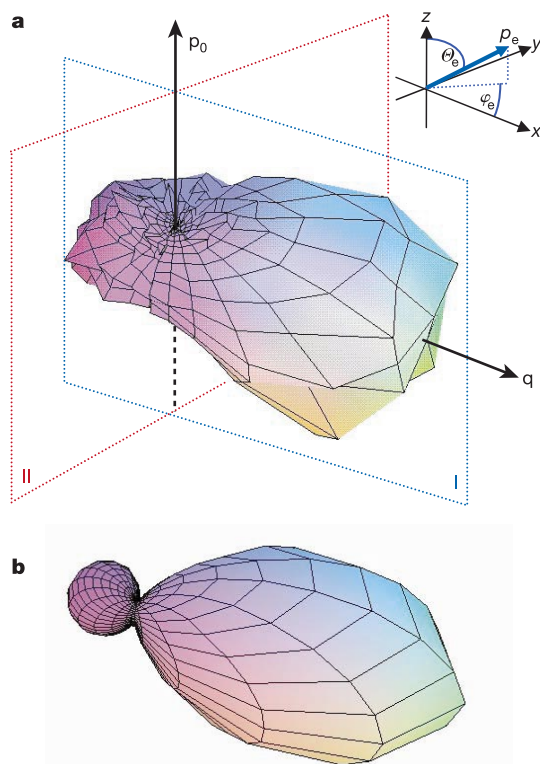


Figure 2 Three-dimensional images of electrons ejected in 100 MeV per a.m.u. $C^{6+} + He$ collisions. Experimental data **a** and theoretical results **b** are shown for an electron energy of $E_e = 6.5$ eV and a momentum transfer of $q = 0.75$ a.u. $\mathbf{p}_0$ and $\mathbf{q}$ are the initial projectile momentum and the momentum transfer. The scattering plane (cut I) and the perpendicular plane (cut II) are indicated by the blue and red lines, respectively. The coordinate system in the upper right corner illustrates the polar (θ_e) and azimuthal (ϕ_e) electron emission angles.

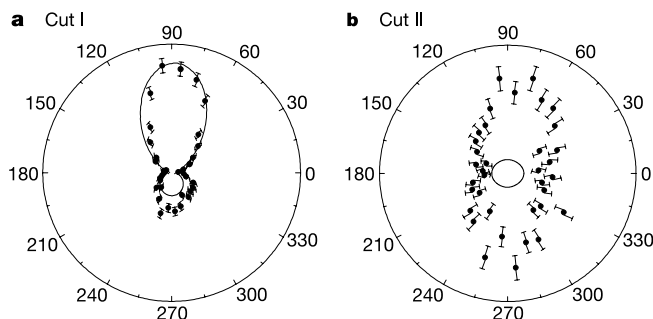


Figure 3 Triply differential single ionization cross-sections for selected electron emission planes. Experimental (solid circles) and theoretical (solid curves) results are shown for the scattering plane (**a**, blue line in Fig. 2) and for the perpendicular plane (**b**, red line in Fig. 2). Orientation in Figs 3 and 4 is given in degrees.

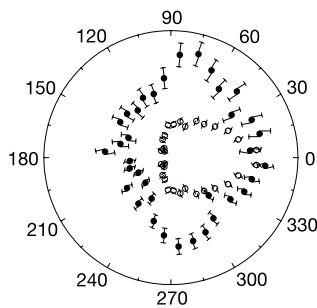


Figure 4 Triply differential single ionization cross-sections for the perpendicular plane for 2 MeV per a.m.u. $C^{6+} + He$ collisions. Open symbols, $E_0 = 4$ eV, $q = 0.7$ a.u.; solid symbols, $E_0 = 1$ eV, $q = 1.5$ a.u.

relative to $\mathbf{q}'$. If this rotation, which does not affect the direction of the electron momentum, is 90° , the electron ends up in the perpendicular plane ($\mathbf{p}_0, \mathbf{q}'$) which started off as the scattering plane after the initial projectile–electron interaction. Because the electron does not take part in the second step, its polar angular distribution remains unchanged leading to a peak at the same angle as in the scattering plane.

This mechanism has an important impact on basic features characteristic to ionization which were previously thought to be understood. What appears to be a recoil peak in the scattering plane emerges as a ‘recoil ring’ in the measured three-dimensional electron emission pattern, which is nearly isotropic with respect to rotation about the projectile beam axis. This behaviour, not reproduced by theory, can also be explained qualitatively by the mechanism suggested above: if the projectile–target nucleus interaction leads to a rotation of $\mathbf{q}$ about the projectile beam axis by 180° then the ionized electron is emitted in the direction of $-\mathbf{q}$ (that is, it looks like a ‘recoil peak electron’), although it was not backscattered by the target nucleus. This could be an important contributor to the recoil peak, which so far has not been discussed in the literature. Such contributions would also explain small but systematic underestimations of the height of the recoil peak by theory both for electron and ion impact^{13,17}. Of course, the projectile–target nucleus interaction can lead to any rotation of $\mathbf{q}$ between 0° and 180° thus giving rise to the observed ring shape.

This process represents a potentially important few-body effect involving all collision products. In principle, it is contained in the final state wavefunction in our calculation. Furthermore, we note that, for the sake of a graphic explanation, our description of this mechanism represents a simplification. The CDW–HF model treats this process more realistically in that here the projectile–electron and projectile–target nucleus interactions do not occur sequentially, but instead simultaneously. However, the final state wavefunction is only asymptotically exact when at least one of the collision products is at a large distance from the other two particles. Its accuracy is not known when all particles are close together. But this process is expected to be important exactly when all collision products are still close together and may therefore be significantly underestimated by the calculation.

Very recently, in a promising theoretical development, the CDW–approach has been generalized to include non-zero magnetic sub-states in the final state wavefunction¹⁴. Such sub-states would be expected to contribute to the cross-sections out of the scattering plane and therefore offer an alternative or complementary explanation to the one proposed above. However, numeric calculations on fully differential cross-sections with this method are not yet available.

Existing theoretical approaches should be tested relative to their capability of reproducing measured three-dimensional electron emission patterns. Such tests may reveal that qualitatively new theoretical concepts need to be developed in order to achieve a better understanding of ionization processes even at large projectile

energies. The data presented here represent experimental progress for such efforts. Obtaining a satisfactory description of atomic few-body processes, in turn, is an important step towards a better understanding of the general few-body problem. □

Received 26 July; accepted 16 December 2002; doi:10.1038/nature01415.

1. Ehrhardt, H., Jung, K., Knoth, G. & Schlemmer, P. Differential cross sections of direct single electron impact ionization. *Z. Phys. D* **1**, 3–32 (1986).
2. Stefani, G., Avaldi, L. & Camilloni, R. On the relevance of the final ionic state to the $(e,2e)$, TDCS. *J. Phys. B* **23**, L227–L231 (1990).
3. Lahmam-Bennani, A. Recent developments and new trends in $(e,2e)$ and $(e,3e)$ studies. *J. Phys. B* **24**, 2401–2442 (1991).
4. Murray, A. J., Woolf, M. B. J. & Read, F. H. Results from symmetric and non-symmetric energy sharing $(e,2e)$ experiments in the perpendicular plane. *J. Phys. B* **25**, 3021–3036 (1992).
5. Röder, J., Ehrhardt, H., Bray, I., Fursa, D. V. & McCarthy, I. E. Absolute triple differential cross-section for electron-impact ionization of helium at 40 eV. *J. Phys. B* **29**, 2103–2114 (1996).
6. Schulz, M. *et al.* Triply differential single ionization cross-sections in coplanar and non-coplanar geometry for fast heavy ion-atom collisions. *J. Phys. B* **34**, L305–L311 (2001).
7. Ehrhardt, H., Schulz, M., Tekaat, T. & Willmann, K. Ionization of helium: angular correlation of the scattered and ejected electrons. *Phys. Rev. Lett.* **22**, 82–92 (1969).
8. Bray, I. & Stelbovics, A. T. Explicit demonstration of the convergence of the close-coupling method for a Coulomb three-body problem. *Phys. Rev. Lett.* **69**, 53–56 (1992).
9. Rescigno, T. N., Baertschy, M., Isaacs, W. A. & McCurdy, C. W. Collisional breakup in a quantum system of three charged particles. *Science* **286**, 2474–2479 (1999).
10. Belkic, D. A quantum theory of ionisation in fast collisions between ions and atomic systems. *J. Phys. B* **11**, 3529–3552 (1978).
11. Crothers, D. S. F. & McCann, J. F. Ionisation of atoms by ion impact. *J. Phys. B* **16**, 3229–3242 (1983).
12. Jones, S. & Madison, D. H. Evidence of initial-state two-center effects for $(e,2e)$ reactions. *Phys. Rev. Lett.* **81**, 2886–2889 (1998).
13. Madison, D., Schulz, M., Jones, S., Foster, M., Moshhammer, R. & Ullrich, J. Comparison of theoretical and absolute experimental fully differential cross sections for ion-atom impact ionization. *J. Phys. B* **15**, 3297–3314 (2002).
14. Crothers, D. S. F. *et al.* Magnetically quantized continuum distorted waves. *Phys. Rev. Lett.* **88**, 053201–053204 (2002).
15. Ullrich, J. *et al.* Recoil-ion, momentum spectroscopy. *J. Phys. B* **30**, 2917–2974 (1997).
16. Whelan, C. T., Allan, R. J., Walters, H. R. J. & Zhang, X. in $(e,2e)$ and Related Processes (eds Whelan, C. T., Walters, H. R. J., Lahmam-Bennani, A. & Ehrhardt, H.) Vol. 414 of *Series C: Mathematical and Physical Sciences* 1–32 (Kluwer, Dordrecht, 1993).
17. Marchalant, P., Whelan, C. T. & Walters, H. R. J. Second-order effects in $(e,2e)$ excitation-ionization of helium to He^+ ($n = 2$). *J. Phys. B* **31**, 1141–1178 (1998).

Acknowledgements We thank W. Schmitt, R. Mann, R. Dörner, T. Weber, Kh. Khayyat, A. Cassimi, L. Adoui, J. P. Grandin and staff members of the CIRIL for their help performing the experiment at the GANIL; thanks to R. E. Olson and J. Fiol for discussions. The support of the Deutsche Forschungsgemeinschaft within the Leibniz-program, by GSI, the European Union, the NSF and by CIRIL (GANIL at Caen) is acknowledged.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials to M.S. (e-mail: schulz@umr.edu).

Ordering and manipulation of the magnetic moments in large-scale superconducting π -loop arrays

Hans Hilgenkamp*, Ariando*, Henk-Jan H. Smilde*, Dave H. A. Blank*, Guus Rijnders*, Horst Rogalla*, John R. Kirtley† & Chang C. Tsuei†

* Faculty of Science and Technology and MESA⁺ Research Institute, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

† IBM T. J. Watson Research Center, Yorktown Heights, New York 10598, USA

The phase of the macroscopic electron-pair wavefunction in a superconductor can vary only by multiples of 2π when going around a closed contour. This results in quantization of magnetic flux, one of the most striking demonstrations of quantum phase coherence in superconductors^{1–3}. By using superconductors with unconventional pairing symmetry^{4–7}, or by incorporating π -Josephson junctions⁸, a phase shift of π can be introduced in

such loops^{7,9,10}. Under appropriate conditions, this phase shift results in doubly degenerate time-reversed ground states, which are characterized by the spontaneous generation of half quanta of magnetic flux, with magnitude $1/2 \Phi_0$ ($\Phi_0 = h/2e = 2.07 \times 10^{-15}$ Wb) (ref. 7). Until now, it has only been possible to generate individual half flux quanta. Here we report the realization of large-scale coupled π -loop arrays based on $\text{YBa}_2\text{Cu}_3\text{O}_7$ -Au-Nb Josephson contacts^{11,12}. Scanning SQUID (superconducting quantum interference device) microscopy has been used to study the ordering of half flux quanta in these structures. The possibility of manipulating the polarities of individual half flux quanta is also demonstrated. These π -loop arrays are of interest as model systems for studying magnetic phenomena—including frustration effects—in Ising antiferromagnets^{13–18}. Furthermore, studies of coupled π -loops can be useful for designing quantum computers based on flux-qubits^{19–23} with viable quantum error correction capabilities^{24,25}.

The half-flux-quantum effect in single π -loops has been used in various phase-sensitive pairing symmetry tests to provide an unambiguous signature for d -wave order parameter pairing symmetry in high-transition-temperature (high- T_c) copper oxide superconductors^{7,9}, and as a potentially practical application of d -wave superconductivity¹⁰. These experiments were mainly conducted by using grain boundaries in high- T_c superconductors, induced by the epitaxial deposition on tri- or tetracrystalline substrates^{7,10} or by using Josephson junctions between a conventional superconductor and $\text{YBa}_2\text{Cu}_3\text{O}_7$ (ref. 9).

However, the techniques used thus far for the fabrication of π -loops do not lend themselves to the construction of coupled π -loop arrays. To this end, we have developed thin-film ramp-type $\text{YBa}_2\text{Cu}_3\text{O}_7$ -Au-Nb Josephson contacts, with high critical current density; the fabrication procedure of these contacts has been described in detail in refs 11 and 12. First, bilayers of 150-nm [001]-oriented $\text{YBa}_2\text{Cu}_3\text{O}_7$ and 100-nm SrTiO_3 are epitaxially grown by pulsed-laser deposition on [001]-oriented SrTiO_3 single-crystal substrates. In these bilayers, the basic layout of the structures that generate half flux quanta, which will be described below, is defined by photolithography and Ar-ion milling. This process results in interfaces with a slope of 15 – 20° with respect to the substrate, which provides access to the a - b planes of the $\text{YBa}_2\text{Cu}_3\text{O}_7$ and allows the exploitation of d -wave phase effects. Special care is taken to align all interfaces accurately along one of the $\text{YBa}_2\text{Cu}_3\text{O}_7$ $\langle 100 \rangle$ axes. After etching the ramp and cleaning the sample, a 7-nm $\text{YBa}_2\text{Cu}_3\text{O}_7$ interlayer is deposited (the function and properties of

which are described in ref. 12), followed by the *in situ* pulsed-laser deposition of a 6-nm Au barrier layer. A 160-nm Nb counter electrode is then formed by d.c. sputtering and structured by lift-off. Subsequently, the redundant, uncovered Au is removed by ion milling.

Each chip contained several reference junctions. At temperature $T = 4.2$ K, these showed a typical critical current per micrometre junction width of $I_c/w \approx 0.1 \text{ mA } \mu\text{m}^{-1}$. From this, a value for the Josephson penetration depth $\lambda_j \approx 1 \mu\text{m}$ ($T = 4.2$ K) is deduced, which is the characteristic length scale over which Josephson vortices (fluxons) extend. The sample magnetic fields were imaged with a high-resolution scanning SQUID microscope^{26–29}. The SQUID microscope images shown here were made at a temperature of 4.2 K, with the sample cooled and imaged in a magnetic induction $< 0.5 \mu\text{T}$. The images of Figs 1 and 2 were made with an octagonal pick-up loop $4 \mu\text{m}$ in diameter; the image of Fig. 3 was made with a square pick-up loop $8 \mu\text{m}$ on a side.

The first configuration for which we investigated the generation and coupling of half-integer flux quanta is the zigzag array¹¹, shown schematically in Fig. 1a inset. In this structure, the d -wave order parameter of the $\text{YBa}_2\text{Cu}_3\text{O}_7$ induces a difference of π in the Josephson phase shift $\Delta\phi$ across the $\text{YBa}_2\text{Cu}_3\text{O}_7$ -Au-Nb barrier for neighbouring facets. For facet lengths a in the wide limit, that is, $a \gg \lambda_j$, the lowest-energy ground state of the system is expected to be characterized by the spontaneous generation of a half-integer flux quantum at each corner. This half-fluxon provides a further π -phase change between neighbouring facets, either adding or subtracting to the d -wave induced π -phase shift, depending on the half-flux-quantum polarity. In both cases, this leads to a lowering of the Josephson coupling energy across the barrier, as this energy is proportional to $(1 - \cos \Delta\phi)$.

In the scanning SQUID microscopy image presented in Fig. 1a, a spontaneously induced magnetic flux is clearly seen at every corner of the zigzag structure. For this sample $a = 40 \mu\text{m}$, which implies that the facets are well within the wide limit. The observed corner fluxons are arranged in an antiferromagnetic fashion. This antiferromagnetic ordering was found to be very robust, occurring for many cool-downs and for different samples with comparable geometries. Deviations from an antiferromagnetic arrangement were only observed when a magnetic field was applied during cool-down, or when an Abrikosov vortex was found trapped in (or near) the junction interface. The total flux per Josephson vortex in the faceted array of Fig. 1a was estimated to be $\Phi = 0.42 \pm 0.12 \Phi_0$ by fitting the observed integrated flux through the SQUID pick-up

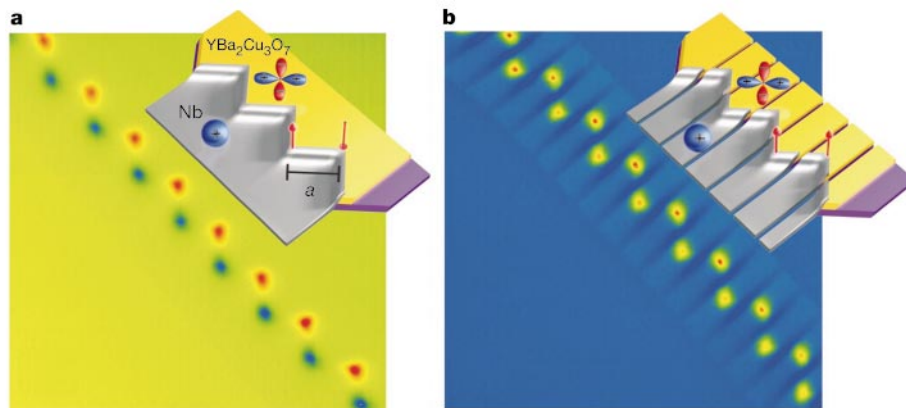


Figure 1 Generation of half flux quanta in connected and unconnected $\text{YBa}_2\text{Cu}_3\text{O}_7$ -Au-Nb zigzag structures. Shown are scanning SQUID micrographs of **a**, 16 antiferromagnetically ordered half-integer flux quanta at the corners of a connected zigzag structure, and **b**, 16 ferromagnetically ordered half flux quanta at the corners of an unconnected zigzag structure ($T = 4.2$ K). The geometries of the structures are shown schematically in the insets. The purple and yellow layers represent the $\text{YBa}_2\text{Cu}_3\text{O}_7$ - SrTiO_3 bilayer; the Nb layer

is indicated in grey. The s -wave and d -wave order parameter symmetries of the Nb and the $\text{YBa}_2\text{Cu}_3\text{O}_7$ are represented by the round and four-leaved-clover-shaped icons, respectively. The facet length a is $40 \mu\text{m}$ in these measurements. In **a**, the yellow–red dots represent flux with opposite polarity to the blue dots, with the zero background flux indicated in green. In **b**, all corner fluxes have the same polarity and are presented by the yellow–red dots, the zero background flux now being indicated in blue.

loop to a model of an array of magnetic monopole sources with alternating signs, localized at the meeting points of the facets⁷. The Josephson vortices were quite uniform, as evidenced by a standard deviation of the observed peak heights of about 0.12 times the average peak height.

In the zigzag configuration, all the half-fluxons are generated in a singly connected superconducting structure; the question therefore arises as to whether the antiferromagnetic ordering is due to a magnetic interaction between the fractional fluxons, or to an interaction via the superconducting connection between the corners. To investigate this, we have also fabricated arrays of corner junctions, in a similar configuration as the zigzag arrays but with 2.5- μm -wide slits etched halfway between the corners. In this situation there is no superconducting connection between the separate flux-generating corner junctions. For a distance between the corners equal to the facet length in the connected array, $a = 40\text{ }\mu\text{m}$, a ferromagnetic arrangement of the fractional flux quanta was observed (Fig. 1b). The magnetic interaction between the half flux quanta at these distances is expected to be very weak, and alignment along minute spurious background fields in the scanning SQUID microscope is anticipated to be the dominating mechanism for their parallel arrangement. When the distance was decreased to about $20\text{ }\mu\text{m}$, with a slit width of $1.5\text{ }\mu\text{m}$, a tendency towards an antiferromagnetic coupling was observed. This was explored further in coupled two-dimensional arrays, which will be described below.

By comparing the behaviour of the connected and the unconnected arrays, we conclude that the antiferromagnetic ordering in the connected zigzag array with $a = 40\text{ }\mu\text{m}$ (Fig. 1a) arises through the phase of the wavefunction in this singly connected superconducting structure. We propose that the basic mechanism for this is the following. During cool-down, the zigzag array undergoes a transition from the small to the wide facet limit, as the Josephson penetration depth is inversely proportional to the square root of the

junctions' critical current density. During this transition the flux is generated, converging ultimately to $1/2\Phi_0$ in the limit $\lambda_j \ll a$. For the intermediate regime, $\lambda_j \approx a$, the flux is smaller—the facet length being too small to accommodate a complete half-fluxon. In that case, to obtain overall the lowest energy state, the phase change associated with the flux generated at one corner is compensated with an equal, but opposite, phase-change at the neighbouring corner, yielding an antiferromagnetic arrangement of the fractional fluxes. When cooled down further, this antiferromagnetic ordering freezes in, and the fractional fluxes grow to complete half-integer flux quanta.

For unconnected corner junctions at sufficiently close distances, the half flux quanta are becoming antiferromagnetically arranged owing to magnetic interaction. To further explore this, we have realized two-dimensional triangular arrays. This geometry is of particular interest, as with a preferential antiferromagnetic coupling between the half-integer flux quanta, it provides a model example of a strongly frustrated system, characterized by a highly degenerate ground state¹³. Until now, such systems have been studied with, for example, arrays of all-low- T_c superconducting rings biased at an external magnetic flux of $1/2\Phi_0$ per ring^{16,17}, and with low- T_c Josephson junction arrays (see, for example, ref. 18 and references therein). Fluctuations in the dimensions of those rings, resulting in variations in the flux bias and a lifting of the degeneracy, have been found to be a major complication in these investigations¹⁶. The spontaneously generated flux in the $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}\text{-Au-Nb}$ structures provides an advantage in this respect, as the two flux states in these elements are intrinsically degenerate.

Figure 2 depicts a scanning SQUID micrograph of the arrangement of the fractional flux quanta in a section of such a triangular array. For this sample, the distance between the half-fluxons is $12\text{ }\mu\text{m}$. In total, about 25,000 half flux quanta were achieved on a single chip, of size $5\text{ mm} \times 10\text{ mm}$.

We verified the predominantly antiferromagnetic nearest-neighbour interaction by determining the bond-order parameter σ , following the method described in refs 16 and 30. In this method, $\sigma = 1 - (n_{\text{AF}}/2x_+x_-)$, with n_{AF} the fraction of antiferromagnetic bonds, and $x_+(x_-)$ the concentrations of half-integer flux quanta pointing upwards (downwards), respectively. For $\sigma < 0$, an additional fraction of antiferromagnetic bonds exists, as compared with the completely random arrangement, which yields $\sigma = 0$. For

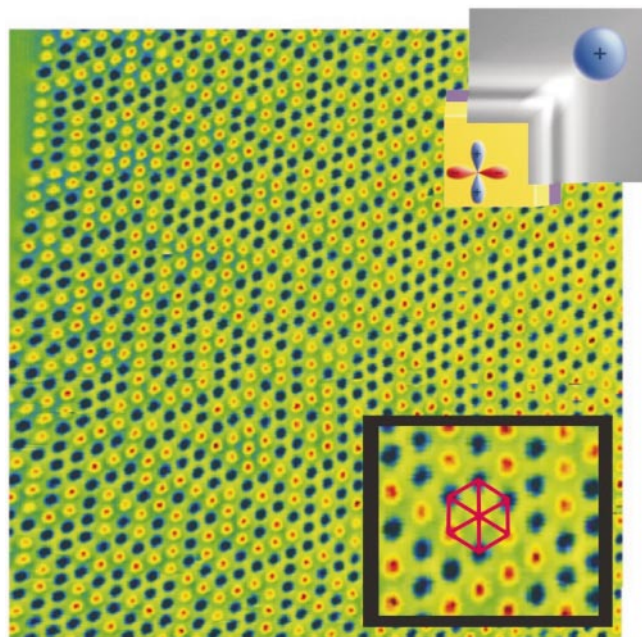


Figure 2 Scanning SQUID micrograph of the flux state in a section of a triangular lattice ($T = 4.2\text{ K}$). Each dot represents a half-fluxon, generated by a corner junction with a facet length of $5\text{ }\mu\text{m}$, shown schematically at top right. All corner junctions are electrically isolated from each other, and the distance between their centre points is $12\text{ }\mu\text{m}$. At bottom right a $4\times$ magnified area is shown, with an overlay of lines forming a body-centred hexagon connecting seven of the half flux quanta, as an illustration of the triangular lattice geometry.

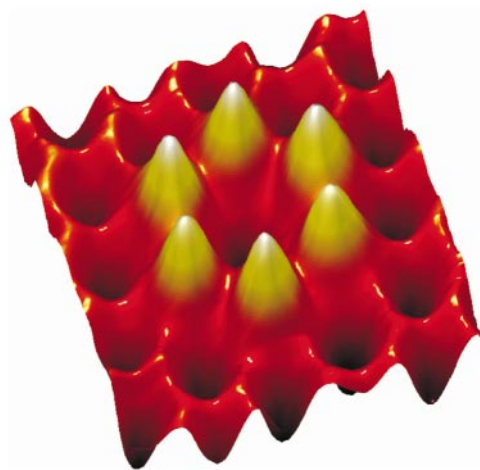


Figure 3 Scanning SQUID micrograph of a two-dimensional triangular array of π -loops ($10\text{-}\mu\text{m}$ -wide facets, $25\text{-}\mu\text{m}$ spacing) in an $80\text{ }\mu\text{m} \times 80\text{ }\mu\text{m}$ area. In this array, a hexagonal pattern of six half flux quanta was written by reversing the polarity of selected elements at zero field and $T = 4.2\text{ K}$, after cooling in a magnetic field so that all half-fluxons initially have the same polarity. In this image, the dark-red valleys represent magnetic flux with opposite polarity to the flux represented by the yellow peaks.

the triangular lattice depicted in Fig. 2, σ was found to be -0.1 , providing evidence for the predominant antiferromagnetic nearest-neighbour coupling of the individual half-integer flux quanta in this system. For the next-nearest-neighbour interaction, at a distance of $\sim 21 \mu\text{m}$, σ was close to 0. It is probable that the pattern of half flux quanta shown in Fig. 2 includes some frozen-in disorder, as with the used cool-down speed of about 10 mK s^{-1} , the large-scale array will not have sufficient time to relax to an overall energetic ground state. This source of disorder may be reduced by lowering the cool-down speed.

The spontaneous currents generated in the elements of the triangular array of Fig. 2 were remarkably uniform, as indicated by a full-width at half-maximum of the distribution in peak amplitudes, which was about 0.28 times the average peak height. We do not believe that this observed spread is a dominant source of disorder in the π -ring array cooling experiments, because it will cause fluctuations in the magnitude, as opposed to fluctuations in the sign, of the ring–ring coupling. A detailed comparison between the ordering in two-dimensional half-fluxon lattices generated in frustrated and unfrustrated geometries—the latter including, for example, square and honeycomb lattices—is under investigation.

Once cooled down to $T = 4.2 \text{ K}$, the energy required to alter the polarity of a half flux quantum is considerably larger than the thermal energy ($E_b \approx I_c \Phi_0 / \pi > 10^4 \text{ K}$), which is reflected in the temporal stability of the flux pattern. But by applying locally a magnetic field, it is possible to manipulate the polarity of the individual half-integer flux quanta, enabling the possibilities of storing information or of constructing desired patterns of fractional flux. This is demonstrated in Fig. 3, showing six half-fluxons in a hexagonal arrangement, with opposite polarity to the surrounding half flux quanta in this triangular lattice, spaced at $25 \mu\text{m}$. The polarity of the half flux quanta in this sample with facet length $a = 10 \mu\text{m}$ was set by scanning a SQUID susceptometer³¹ over the junctions at a rate of 0.05 mm s^{-1} while applying a 2-mA current through the excitation coil (this coil is $21 \mu\text{m}$ in diameter, and concentric and co-planar with the $8 \mu\text{m}$ square SQUID pick-up loop). This corresponds to a field of approximately $50 \mu\text{T}$ at the junction. Reversing the current direction reversed the resultant polarity of the half-vortex. The critical field required to flip the polarity of the half-vortex is in rough agreement with numerical calculations³². We expect that, apart from the spontaneous flux formation, the writing of half flux quanta patterns will provide a diverse basis for both fundamental studies and potential applications. □

Received 14 October 2002; accepted 14 January 2003; doi:10.1038/nature01442.

1. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
2. Deaver, B. S. Jr & Fairbank, W. M. Experimental evidence for quantized flux in superconducting cylinders. *Phys. Rev. Lett.* **7**, 43–46 (1961).
3. Doll, R. & Näbauer, M. Experimental proof of magnetic flux quantization in a superconducting ring. *Phys. Rev. Lett.* **7**, 51–52 (1961).
4. Geshkenbein, V. B. & Larkin, A. I. The Josephson effect in superconductors with heavy fermions. *JETP Lett.* **43**, 395–399 (1986).
5. Geshkenbein, V. B., Larkin, A. I. & Barone, A. Vortices with half magnetic flux quanta in “heavy-fermion” superconductors. *Phys. Rev. B* **36**, 235–238 (1987).
6. Sigrist, M. & Rice, T. M. Paramagnetic effect in high- T_c superconductors—a hint for d -wave superconductivity. *J. Phys. Soc. Jpn* **61**, 4283–4286 (1992).
7. Tsuei, C. C. & Kirtley, J. R. Pairing symmetry in cuprate superconductors. *Rev. Mod. Phys.* **72**, 969–1016 (2000).
8. Bulaeviskii, L. N., Kuzii, V. V. & Sobyannin, A. A. Superconducting system with weak coupling to the current in the ground state. *JETP Lett.* **25**, 290–294 (1977).
9. Van Harlingen, D. J. Phase-sensitive tests of the symmetry of the pairing state in the high-temperature superconductors—Evidence for $d_{x^2-y^2}$ symmetry. *Rev. Mod. Phys.* **67**, 515–535 (1995).
10. Schulz, R. R. *et al.* Design and realization of an all d -wave dc π -superconducting quantum interference device. *Appl. Phys. Lett.* **76**, 912–914 (2000).
11. Smilde, H. J. H. *et al.* d -Wave-induced Josephson current counterflow in $\text{YBa}_2\text{Cu}_3\text{O}_7/\text{Nb}$ zigzag junctions. *Phys. Rev. Lett.* **88**, 057004 (2002).
12. Smilde, H. J. H., Hilgenkamp, H., Rijnders, G., Rogalla, H. & Blank, D. H. A. Enhanced transparency ramp-type Josephson contacts through interlayer deposition. *Appl. Phys. Lett.* **80**, 4579–4581 (2002).
13. Aeppli, G. & Chandra, P. Seeking a simple complex system. *Science* **275**, 177–178 (1997).
14. Moessner, R. & Sondhi, S. L. Ising models of quantum frustration. *Phys. Rev. B* **63**, 224401 (2001).

15. Chandra, P. & Doucot, B. Possible spin-liquid state at large S for the frustrated square Heisenberg lattice. *Phys. Rev. B* **38**, 9335–9338 (1988).
16. Davidovic, D. *et al.* Magnetic correlations, geometrical frustration, and tunable disorder in arrays of superconducting rings. *Phys. Rev. B* **55**, 6518–6540 (1997).
17. Pannetier, B., Chaussy, J., Rammal, R. & Villegier, J. C. Experimental fine tuning of frustration: Two-dimensional superconducting networks in a magnetic field. *Phys. Rev. Lett.* **53**, 1845–1848 (1984).
18. Lerch, Ph., Leemann, Ch., Theron, R. & Martinoli, P. Dynamics of the phase transition in proximity-effect arrays of Josephson junctions at full frustration. *Phys. Rev. B* **41**, 11579–11582 (1990).
19. Mooij, J. E. *et al.* Josephson persistent-current qubit. *Science* **285**, 1036–1039 (1999).
20. Van der Wal, C. *et al.* Quantum superposition of macroscopic persistent-current states. *Science* **290**, 773–777 (2000).
21. Friedman, J. R., Patel, V., Chen, W., Tolpygo, S. K. & Lukens, J. E. Quantum superposition of distinct macroscopic states. *Nature* **406**, 43–46 (2000).
22. Ioffe, L. B., Geshkenbein, V. B., Feigel'man, M. V., Fauchère, A. L. & Blatter, G. Environmentally decoupled sds-wave Josephson junctions for quantum computing. *Nature* **398**, 679–681 (1999).
23. Blais, A. & Zagoskin, A. Operation of universal gates in a solid-state quantum computer based on clean Josephson junctions between d -wave superconductors. *Phys. Rev. A* **61**, 042308 (2000).
24. Preskill, J. Reliable quantum computers. *Proc. R. Soc. Lond. A* **454**, 385–410 (1998).
25. Bennett, C. H. & DiVincenzo, D. P. Quantum information and computation. *Nature* **404**, 247–255 (2000).
26. Rogers, F. P. *A Device for Experimental Observation of Flux Vortices Trapped in Superconducting Thin Films*. Thesis, MIT, Boston (1983).
27. Vu, L. N. & Van Harlingen, D. J. Design and implementation of a scanning SQUID microscope. *IEEE Trans. Appl. Supercond.* **3**, 1918–1921 (1993).
28. Black, R. C. Magnetic microscopy using a liquid nitrogen cooled $\text{YBa}_2\text{Cu}_3\text{O}_7$ superconducting quantum interference device. *Appl. Phys. Lett.* **62**, 2128–2130 (1993).
29. Kirtley, J. R. *et al.* High-resolution scanning SQUID microscope. *Appl. Phys. Lett.* **66**, 1138–1140 (1995).
30. Cowley, J. M. An approximate theory of order in alloys. *Phys. Rev.* **77**, 669–675 (1950).
31. Gardner, B. W. *et al.* Scanning superconducting quantum interference device susceptometry. *Rev. Sci. Instrum.* **72**, 2361–2364 (2001).
32. Kirtley, J. R., Moler, K. A. & Scalapino, D. J. Spontaneous flux and magnetic-interference patterns in $0-\pi$ Josephson junctions. *Phys. Rev. B* **56**, 886–891 (1997).

Acknowledgements We thank R. H. Koch, J. Mannhart, D. M. Newns and D. J. Scalapino for discussions. This work was supported by the Dutch Foundation for Research on Matter (FOM), the Netherlands Organization for Scientific Research (NWO), the Royal Dutch Academy of Arts and Sciences (KNAW) and the European Science Foundation (ESF) PiShift programme.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.H. (e-mail: H.Hilgenkamp@utwente.nl).

Superconductivity in two-dimensional CoO_2 layers

Kazunori Takada^{*,‡}, Hiroya Sakurai[†], Eiji Takayama-Muromachi[†], Fujio Izumi^{*}, Ruben A. Dilanian^{*} & Takayoshi Sasaki^{*,‡}

^{*} Advanced Materials Laboratory, National Institute for Materials Science, Tsukuba, Ibaraki 305-0044, Japan

[†] Superconducting Materials Center, National Institute for Materials Science, Tsukuba, Ibaraki 305-0044, Japan

[‡] CREST, Japan Science and Technology Corporation

Since the discovery of high-transition-temperature (high- T_c) superconductivity in layered copper oxides¹, many researchers have searched for similar behaviour in other layered metal oxides involving 3d-transition metals, such as cobalt and nickel. Such attempts have so far failed, with the result that the copper oxide layer is thought to be essential for superconductivity. Here we report that $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ ($x \approx 0.35$, $y \approx 1.3$) is a superconductor with a T_c of about 5 K. This compound consists of two-dimensional CoO_2 layers separated by a thick insulating layer of Na^+ ions and H_2O molecules. There is a marked resemblance in superconducting properties between the present material and high- T_c copper oxides, suggesting that the two systems have similar underlying physics.

The $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ sample was obtained through a chemical

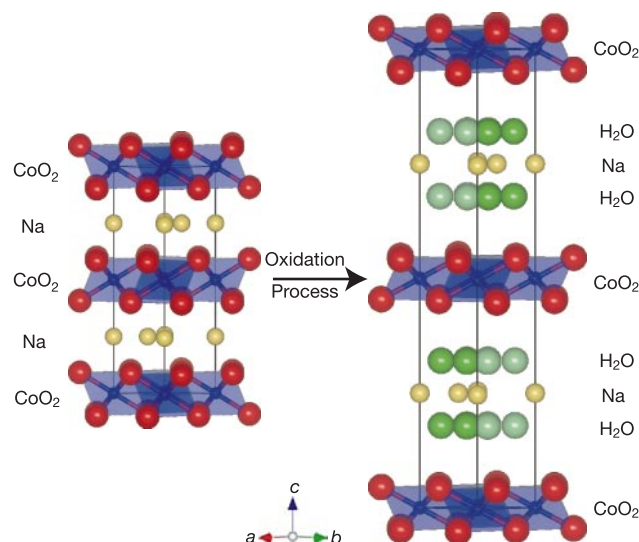


Figure 1 Structural views of $\text{Na}_{0.7}\text{CoO}_2$ (left) and $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ (right), where Na and H_2O sites are partially occupied. $\text{Na}_{0.7}\text{CoO}_2$ was prepared from Na_2CO_3 (99.99%) and Co_2O_3 (99.9%) by solid-state reaction at 800 °C for 8 h under oxygen gas flow. A fivefold excess of Br_2 regarding the Na content was dissolved in acetonitrile (CH_3CN). A well pulverized powder of $\text{Na}_{0.7}\text{CoO}_2$ was immersed in the $\text{Br}_2/\text{CH}_3\text{CN}$ solution for 5 days to deintercalate Na^+ ions; then the product was filtered, washed with CH_3CN and distilled water, and finally dried in an ambient atmosphere. No impurities were detected by energy dispersive X-ray analysis (EDX).

oxidation process (hereafter referred to as the ‘oxidation process’) from a parent compound of $\text{Na}_{0.7}\text{CoO}_2$ (ref. 2). Before the oxidation process, $\text{Na}_{0.7}\text{CoO}_2$ has a hexagonal layered structure (space group $P6_3/mmc$) consisting of the two-dimensional (2D) layers of CoO_2 and charge-balancing Na^+ ions (see the left side of Fig. 1)³. Figure 2 shows an X-ray diffraction (XRD) pattern of the product obtained through the oxidation process. All the reflections of the product

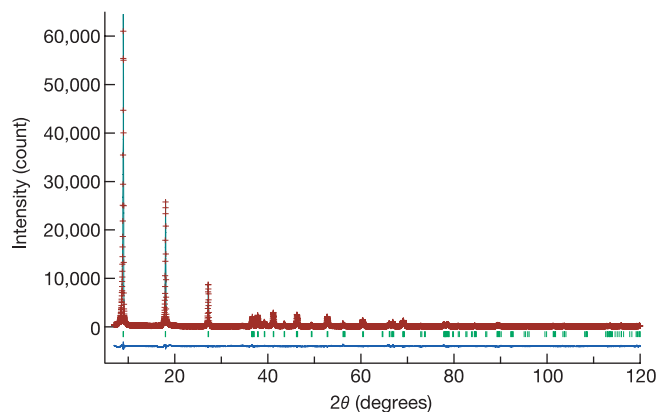


Figure 2 Rietveld refinement patterns for $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$. The observed diffraction intensities are represented by red plus signs, and the calculated pattern by the green solid line. The blue curve at the bottom represents the weighted difference, $(Y_{io} - Y_{ic})/\sigma$, where Y_{io} , Y_{ic} and σ are the observed and calculated intensities and the statistical uncertainty of the i th point, respectively. The differences were multiplied by 20 to make the plots easier to read. Short light green vertical bars below the observed and calculated patterns indicate the positions of allowed Bragg reflections. Powder XRD was performed by an X-ray diffractometer (RINT-2100S, Rigaku) using $\text{CuK}\alpha$ radiation. The oxidation process slightly decreased a from 2.8292(3) Å for the parent compound to 2.8230(2) Å, where the parentheses indicate standard deviations, probably because Co^{3+} ions were partially oxidized to smaller Co^{4+} ions. On the other hand, c increased dramatically from 10.9628(8) Å to 19.6207(14) Å.

Table 1 Structure parameters of $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$

Atom	Site	x	y	z	g	U (Å ²)
Co	2a	0	0	0	1	0.0063(6)
O	4f	1/3	2/3	0.0451(3)	1	0.0162(12)
Na1	2d	2/3	1/3	1/4	0.159(4)	$= U(\text{O})$
Na2	2b	0	0	1/4	0.192*	$= U(\text{O})$
WO1†	12k	0.174(13)	$= 2x(\text{WO1})$	0.1793(2)	0.070(12)	$= U(\text{O})$
WO2†	12k	0.370(12)	$= x(\text{WO2})/2$	$= z(\text{WO1})$	0.176(12)	$= U(\text{O})$

Space group: $P6_3/mmc$; $a = 2.8230(2)$ Å and $c = 19.6207(14)$ Å; $R_{\text{wp}} = 11.58\%$ ($\chi^2 = 4.70$), $R_F = 2.84\%$ and $R_E = 3.53\%$. g , occupancy; U , isotropic displacement parameter; x , y and z , fractional coordinates.

*The total Na content x was fixed at 0.351, which was determined by ICP-AES.

†WO denotes an H_2O molecule, whose atomic scattering factors were set equal to the sums of those of one O and two H atoms.

were indexable on the basis of space group $P6_3/mmc$. A marked increase in c was observed, which supports the idea that the intercalation of certain ‘guest’ molecules occurred in the oxidation process in addition to the deintercalation of Na^+ ions. H_2O is the most probable candidate for the guest. The composition calculated on this assumption from Na and Co contents determined by inductively coupled plasma atomic-emission spectrometry (ICP-AES) was $\text{Na}_{0.35}\text{CoO}_2 \cdot 1.3\text{H}_2\text{O}$. In addition, the analysis of species evolved from heated samples by infrared spectroscopy showed that the H_2O content was 1.24 per formula unit.

Next, we carried out Rietveld analysis⁴ of the XRD data, adopting a structural model shown in the right side of Fig. 1. In this model, H_2O molecules were located at 12k sites on the basis of electron-density distribution obtained by a maximum-entropy method⁴. Table 1 lists the final structure parameters, and Fig. 2 shows Rietveld refinement patterns. The resulting H_2O content was 1.47 per formula unit, which is close to the analytical values. The final R factors were fairly low, which provides evidence for the validity of our present structural model.

The magnetic susceptibility (χ) of $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ is plotted in Fig. 3 as a function of temperature. In the measurements under the external magnetic field $H = 20$ Oe, a steep decrease of susceptibility was observed at about 5 K both in the zero-field cooling and field cooling processes. The χ at 2 K in the field cooling process was -1.6×10^{-3} e.m.u. g^{-1} , which was 6.5% of the theoretical value for perfect diamagnetism. Twice-as-large diamagnetism ($\chi = -3.4 \times 10^{-3}$ e.m.u. g^{-1} , 13% of the theoretical value) was

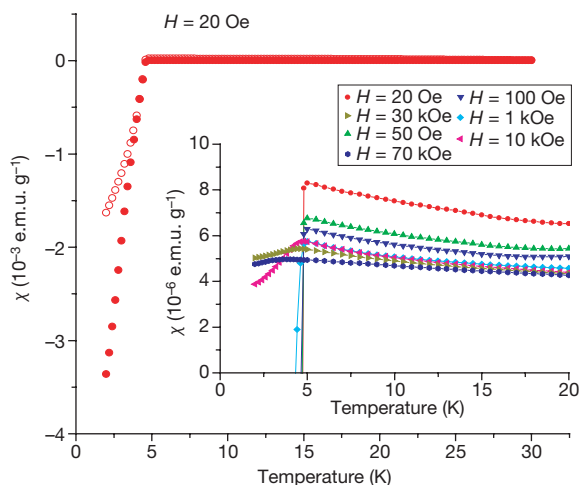


Figure 3 Magnetic susceptibility (χ) of $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$. Filled circles, zero-field cooling; open circles, field cooling. The susceptibility was measured in a magnetic field of 20 Oe. The inset shows χ measured under various magnetic fields by the zero-field cooling process. The magnetization was measured by a superconducting quantum interference device (SQUID) magnetometer (MPMS XL, Quantum Design).

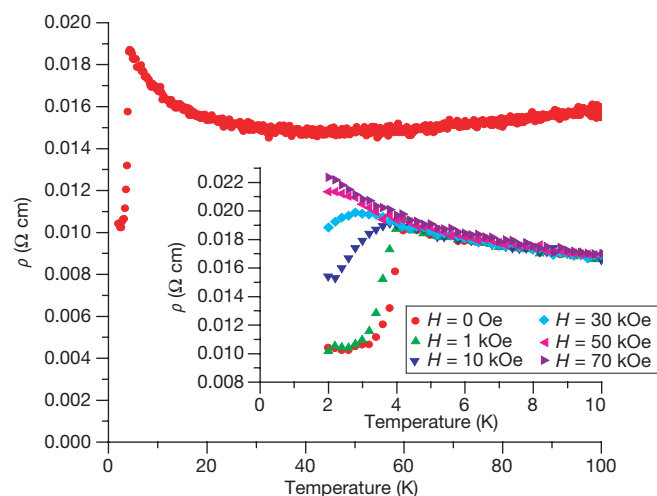


Figure 4 Resistivity (ρ) of $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ under zero magnetic field. The inset figure shows the resistivity measured under various magnetic fields. The electric resistivity was measured on a compressed powder sample by a standard four-probe method using a physical property measurement system (PP700C1, Quantum Design).

observed in the zero-field cooling process. We conclude that the present phase undergoes superconducting transition at about 5 K, because only superconductivity can account for such a large diamagnetism. With increasing H to 1 kOe, the onset temperature of superconductivity (onset T_c) did not change noticeably. With further increase of H up to 70 kOe, downturn of the susceptibility was still observed near 5 K, but the superconducting transition became quite broad and susceptibility was not negative even at 2 K. The M (magnetization)– H curve at 2 K, which is not shown here, indicated that the present material is a superconductor of the second kind with a lower critical field of around 100 Oe. The normal-state susceptibility depended on temperature, and decreased with increasing external magnetic field. $\text{Na}_{0.5}\text{CoO}_2$ shows Curie–Weiss-like magnetic behaviour rather than Pauli paramagnetic behaviour⁵. This suggests that the aforementioned normal-state magnetism reflects the intrinsic nature of the CoO_2 layer, although there is a possibility that the present sample contained a trace amount of magnetic impurity.

The electric resistivity (ρ) of $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$ is shown in Fig. 4. For the present system, it was impossible to prepare a tightly sintered ceramic specimen for the resistivity measurement. Instead, we measured a specimen obtained by room-temperature compression, and zero resistivity was not attained down to 2 K owing to this limitation. Nevertheless, a sharp decrease of resistivity was observed at around 4 K, supporting the superconducting transition. The onset temperature of 4 K is slightly lower than the value determined by the magnetic measurement, and this difference may be due to the variation of water content in the sample. The H_2O content was dependent on the humidity in the atmosphere, and superconductivity was greatly suppressed in a low H_2O -content sample. The superconducting transition became broad with an increasing magnetic field, as shown in Fig. 4, and no decline of resistivity was observed above 50 kOe. This presumably reflects the weak link nature of the specimen. It is, however, worth noting that the onset T_c does not appear to depend significantly on the magnetic field. For instance, onset T_c at 10 kOe was very close to that under zero field.

There is a marked resemblance between the present cobalt oxide and high- T_c copper oxides. In the high- T_c copper oxides, superconductivity occurs in the CuO_2 square lattice. The Cu^{2+} ($S = 1/2$) moments are antiferromagnetically ordered in the CuO_2 plane. With a low level of carrier (hole or electron) doping, the antiferro-

magnetism is suppressed drastically, and the system becomes metallic, followed by the appearance of superconductivity. By analogy, the present material may be understood to be an electron-doped system for a low-spin Co^{4+} ($S = 1/2$) lattice with an electron density (x in $\text{Na}_x\text{CoO}_2 \cdot y\text{H}_2\text{O}$) of around 0.35 per Co atom. As shown in Figs 3 and 4, the increasing external magnetic field broadened the superconducting transition but did not markedly lower the onset T_c . Similar behaviour is always observed in high- T_c copper oxides.

The strong 2D character of the layered copper oxides is believed to be important in superconductivity. In the same way, the large separation of the CoO_2 layers by the introduction of H_2O molecules seems to be essential for inducing superconductivity in the present material (superconductivity is greatly suppressed by the removal of water, accompanied by a shrinkage of the c axis). The main difference between the two systems is that Co ions form a triangular lattice with magnetically frustrated geometry in contrast to the square lattice of the CuO_2 plane. Further studies are needed to elucidate the mechanism of superconductivity. □

Received 18 October 2002; accepted 28 January 2003; doi:10.1038/nature01450.

1. Bednorz, J. G. & Müller, K. A. Possible high T_c superconductivity in the Ba-La-Cu-O system. *Z. Phys. B* **64**, 189–193 (1986).
2. Fouassier, C., Matejka, G., Reau, J.-M. & Hagenmüller, P. Sur de nouveaux bronzes oxygènes de formule Na_xCoO_2 ($x \leq 1$). Le système cobalt-oxygène-sodium. *J. Solid State Chem.* **6**, 532–537 (1973).
3. Balsys, R. J. & Davis, R. L. Refinement of the structure of $\text{Na}_{0.74}\text{CoO}_2$ using neutron powder diffraction. *Solid State Ionics* **93**, 279–282 (1996).
4. Izumi, F. & Ikeda, T. A Rietveld-analysis program RIETAN-98 and its applications to zeolites. *Mater. Sci. Forum* **321–324**, 198–203 (2000).
5. Tanaka, T., Nakamura, S. & Iida, S. Observation of distinct metallic conductivity in NaCo_2O_4 . *Jpn J. Appl. Phys.* **33**, L581–L582 (1994).

Acknowledgements We thank Y. Yajima, S. Takenouchi and H. Komori of our institute for IR, ICP-AES and EDX analysis.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.T. (e-mail: takada.kazunori@nims.go.jp).

..... Dynamical coupling of wind and ocean waves through wave-induced air flow

T. S. Hristov*, S. D. Miller† & C. A. Friehe†‡

* Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland 21218, USA

† Department of Earth System Science, ‡ Department of Mechanical and Aerospace Engineering, University of California, Irvine, California 92697, USA

Understanding the physical mechanisms behind the generation of ocean waves by wind has been a longstanding challenge^{1,2}. Previous studies^{3–6} have assumed that ocean waves induce fluctuations in velocity and pressure of the overlying air that are synchronized with the waves, and numerical models have supported this assumption⁷. In a complex feedback, these fluctuations provide the energy for wave generation. The spatial and temporal structure of the wave-induced airflow therefore holds the key to the physics of wind–wave coupling, but detailed observations have proved difficult. Here we present an analysis of wind velocities and ocean surface elevations observed over the open ocean. We use a linear filter⁸ to identify the wave-induced

air flow from the measurements and find that its structure is in agreement with 'critical-layer' theory³. Considering that the wave-induced momentum flux is then controlled by the wave spectrum and that it varies considerably in vertical direction, a simple parameterization of the total air–sea momentum flux is unlikely to exist.

Over the ocean, turbulent pressure fluctuations in the air initiate small ripples on the water surface⁹ that further grow and develop into wind-driven waves through a poorly understood mechanism. Once present, the surface waves profoundly affect the exchange of momentum and kinetic energy at the air–sea interface. Although experiments have detected strong variability of the interfacial fluxes with sea state, both a clear phenomenological description of the fluxes and insight into the mechanism of exchange have remained elusive^{10–12}. Incomplete understanding of the air–sea interaction reduces the predictive power of climate models as well as of weather and wave forecasts.

The concept of resonant wind–wave interaction assumes that the waves induce velocity $\tilde{\mathbf{v}}$ and pressure $\tilde{p}$ fluctuations in the wind, that in a broad sense are synchronized with the waves. Considering collinear wind and waves, the wind velocity $\mathbf{v}$ is a sum of turbulent fluctuations $\mathbf{v}' \equiv (u', v', w')$ and wave effects $\tilde{\mathbf{v}} \equiv (\tilde{u}, 0, \tilde{w})$, both superimposed on a long-term mean horizontal velocity $\mathbf{U} = (U, 0, 0)$ so that $\mathbf{v} = \mathbf{U} + \mathbf{v}' + \tilde{\mathbf{v}}$. The same decomposition applies to the pressure: $p = P + p' + \tilde{p}$. When the perturbed pressure $\tilde{p}$ is lower on the lee side and higher on the windward side of the wave, the perturbation delivers energy from the air flow to the waves at a

rate $\tilde{E} = \langle \tilde{p}_0 \tilde{\eta} \rangle$ ($\tilde{p}_0$ is the pressure and $\tilde{\eta}$ is the velocity at the interface). The spatio-temporal structure of the wave-induced flow uniquely describes the mechanism of coupling, determines the air pressure distribution on the interface¹³, and therefore the wave growth rate. Agreement between theoretically derived and experimentally observed flow structures would present a conclusive validation for the theory predicting the flow structure. In contrast, because the interfacial pressure distribution does not uniquely determine the overlying flow structure, agreement between measured and predicted wave growth rates only is insufficient to confirm experimentally a wind–wave interaction model¹⁴.

The critical-layer theory³ builds on the concept of resonant wind–wave interaction. There, the turbulent atmospheric flow is considered to be of constant stress $\tau(z) \equiv -\rho \langle u' w' \rangle = \rho u_*^2 = \text{Const.}$, so the mean-wind vertical profile is logarithmic¹⁵ $U(z) = (u_* / \kappa) \log(z/z_0)$. (Here, ρ is the air density, u_* is the friction velocity, z_0 is the ocean surface roughness scale, κ is the von Karman constant, and z is the distance from the unperturbed air–water interface.) The surface waves are generated as growing perturbations of the mean flow $U(z)$, while no wave–turbulence interaction is assumed to occur (quasi-laminar approximation). Because of the large Reynolds number of the wind over the ocean, viscosity is neglected. Within these assumptions the perturbation stream function $\psi(z)$ satisfies the Rayleigh equation^{3,16}:

$$\psi'' - k^2 \psi - U''(U - c)^{-1} \psi = 0 \quad (1)$$

Here $c = (g/k)^{1/2}$ is the phase speed of the wave mode $\eta = A e^{ik(x-ct)}$, k is its wavenumber and g is the acceleration of gravity. The parameter controlling the solution of equation (1), and therefore the wind–wave interaction, is the ratio c/u_* . For equation (1) the height z_c , where $U(z_c) = c$, known as the critical height, is a point of regular singularity. Velocity and pressure fields, obtained from solving equation (1), result in these wind-to-wave fluxes of energy

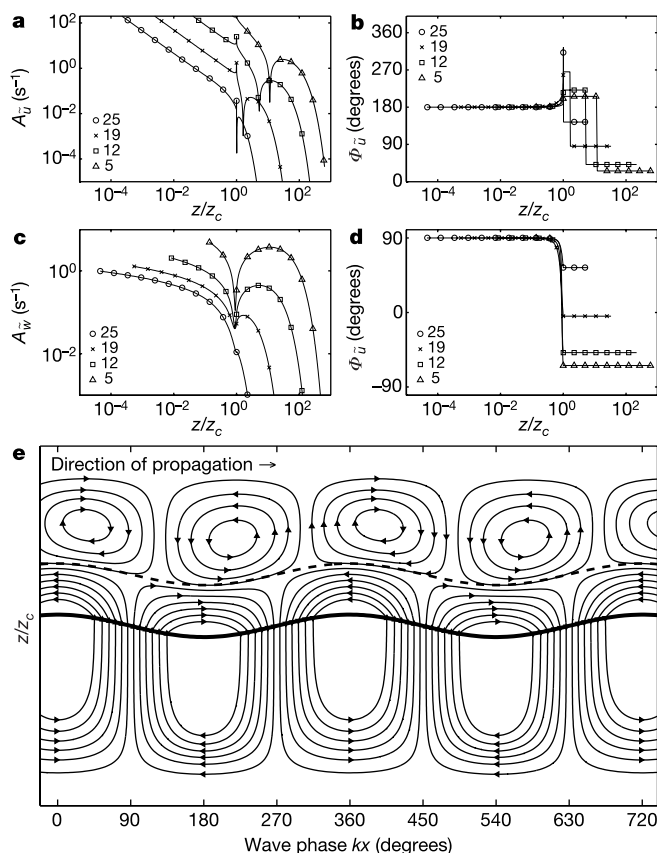


Figure 1 Numerical solutions for the Rayleigh equation (1). **a**, **b**, Amplitude $A_{\tilde{u}}$ (**a**) and phase $\Phi_{\tilde{u}}$ (**b**) of the horizontal wave-induced velocity fluctuations $\tilde{u}(z/z_c)$. **c**, **d**, Amplitude $A_{\tilde{w}}$ (**c**) and phase $\Phi_{\tilde{w}}$ (**d**) of the vertical wave-induced velocity fluctuations $\tilde{w}(z/z_c)$. Each curve corresponds to a value of the parameter c/u_* , as indicated in the symbol key. The flow streamlines (**e**) have been reconstructed from the solution ψ for $c/u_* = 5.75$. The solid line represents the air–water interface and the dashed line is the critical height. The whole eddy structure is propagating to the right following the wave. k is wavenumber.

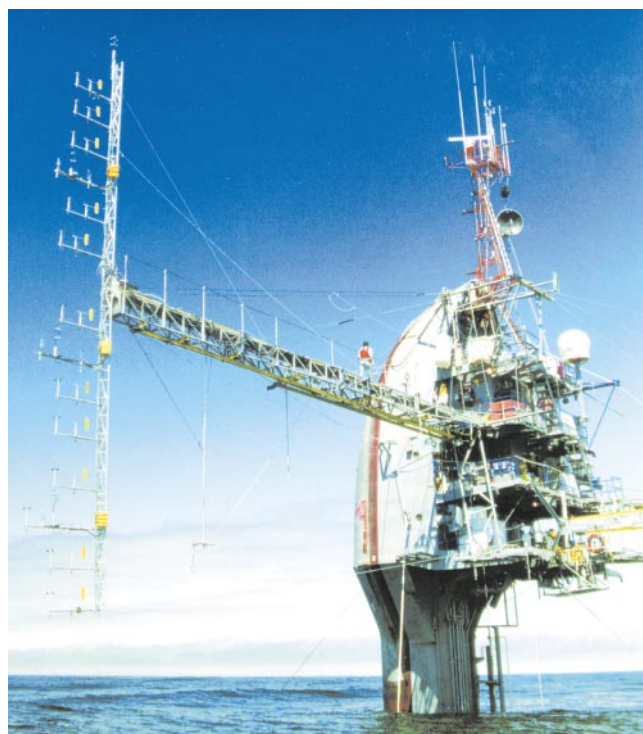


Figure 2 Floating Instrument Platform FLIP moored during the experiment. Four ultrasonic anemometers measured the wind velocity at different heights above the ocean surface and 12 cup anemometers and vanes registered the horizontal wind speed and direction. The sea surface elevation was measured beneath the mast.

$\tilde{E} = -(\pi/2)(u_*/\kappa)^2 \rho k c |\eta|^2 |\psi(z_c)|^2 U''(z_c)/U'(z_c)$, and momentum $\tilde{\tau} \equiv -\rho \langle \tilde{u} \tilde{w} \rangle = \tilde{E}/c$. The dependence of $\tilde{E}$ and $\tilde{\tau}$ on quantities at the critical height z_c indicates that wind-wave energy and momentum exchanges are formally occurring in a thin layer about z_c , known as the critical layer^{3,16}. Away from the critical height the singular term $U''\psi/(U-c)$ vanishes and the solutions are decaying exponents scaled by the wavelength $\psi \propto e^{-kz}$. Thus short (slow) wave signatures in the wind will be significant only close to the surface, while only fluctuations induced by the long (fast) waves will propagate further up.

The complex-valued solution $\psi(z, c/u_*)$ of equation (1) carries information about the wave-induced along-wind $\tilde{u} = -(u_*/\kappa) \times (d\psi/dz)\eta$ and vertical $\tilde{w} = ik(u_*/\kappa)\psi\eta$ velocity. Their normalized amplitudes $A_{\tilde{u}} = |\tilde{u}|/\eta$, $A_{\tilde{w}} = |\tilde{w}|/\eta$ and phases $\Phi_{\tilde{u}} = \arg(\tilde{u}/\eta)$, $\Phi_{\tilde{w}} = \arg(\tilde{w}/\eta)$ appear in Fig. 1. The solution predicts distinct features for the wave-induced flow at z_c : the phase $\Phi_{\tilde{w}}(z)$ is constant on both sides of z_c and discontinuous at z_c (Fig. 1d) and the amplitude $A_{\tilde{w}}(z)$ has a minimum near that point (Fig. 1c). The behaviour of $\tilde{u}$ includes a pair of abrupt changes in amplitude (Fig. 1a) and phase (Fig. 1b) at and above z_c . We will seek to recognize these distinct

features of the wave-induced flow in our field experiment data. The spatial configuration of the wave-induced flow (Fig. 1e), reconstructed from the solutions of equation (1) (Fig. 1a–d), shows two eddy structures brought into contact at the critical height and the streamlines continuous through the air–water interface. A Fourier superposition of flows like the one in Fig. 1e forms the actual wave-induced flow over the sea.

Here we analyse data collected continuously over five days in May 1995 over the open ocean 50 km off Monterey, California, at water depth of 1,500 m from the stable research platform *FLIP* (Fig. 2). From a mast we measured the turbulent wind velocity with four ultrasonic anemometers (accuracy 3% of the instantaneous velocity) at fixed heights of 3.5, 8.7, 13.8 and 18.1 m above the mean ocean level, as well as the surface elevation η (accuracy, 1 cm) directly beneath the mast. The signals were sampled at 50 Hz. The slight motion of the platform was recorded and later removed from the measured wind velocities. During the experiment the conditions ranged from calm, with wind velocity at 10 m height $U_{10} = 2 \text{ m s}^{-1}$, to stormy, with $U_{10} = 14 \text{ m s}^{-1}$. Waves aligned with the wind developed in response to the increasing wind speed and maximum wave height reached 3 m. During the experiment, both u_* and the wave spectrum varied, providing a range of the controlling parameter c/u_* .

The wave-induced flow $\tilde{\mathbf{v}}$ in the wind has long been difficult to detect because turbulence $\mathbf{v}'$ dominates and blurs any wave signature there^{17,18}. A robust and optimal method to separate $\mathbf{v}'$ and $\tilde{\mathbf{v}}$ has been proposed⁸. The linearity of the Rayleigh equation (1) leads to the wave-induced velocity $\tilde{\mathbf{v}}$ being related to the wave elevation η by a linear transform, that is, $\tilde{\mathbf{v}}$ is coherent with η . Therefore, a filter separating $\tilde{\mathbf{v}}$ from $\mathbf{v}'$ must also ensure that property of $\tilde{\mathbf{v}}$. As an estimate for $\tilde{\mathbf{v}}$, we choose the orthogonal projection of the wind signal $\mathbf{v}$ onto the Hilbert space of all the wave-coherent signals. Such an estimate is optimal in the least-squares sense⁸.

Measurements of wind velocity were conducted at fixed heights and over a time period (90 minutes, in this case), they produce a single value of u_* , while the spectrum of waves provides a range of values for c . Therefore, comparison with our experiment requires consideration of the behaviour of $\tilde{u}(z, c/u_*)$ and $\tilde{w}(z, c/u_*)$ at fixed z and u_* and variable c . Introducing $S_{\alpha\beta}$ as the cross-spectral density of the time-series $\alpha(t)$ and $\beta(t)$ and the transfer functions $T_{\tilde{u}}(c) = S_{\tilde{u}\eta}/S_{\eta\eta}$ and $T_{\tilde{w}}(c) = S_{\tilde{w}\eta}/S_{\eta\eta}$, we estimate the normalized amplitudes ($A_{\tilde{u}} = |T_{\tilde{u}}|$ and $A_{\tilde{w}} = |T_{\tilde{w}}|$) and phases ($\Phi_{\tilde{u}} = \arg T_{\tilde{u}}$ and $\Phi_{\tilde{w}} = \arg T_{\tilde{w}}$) of the measured $\tilde{u}$ and $\tilde{w}$.

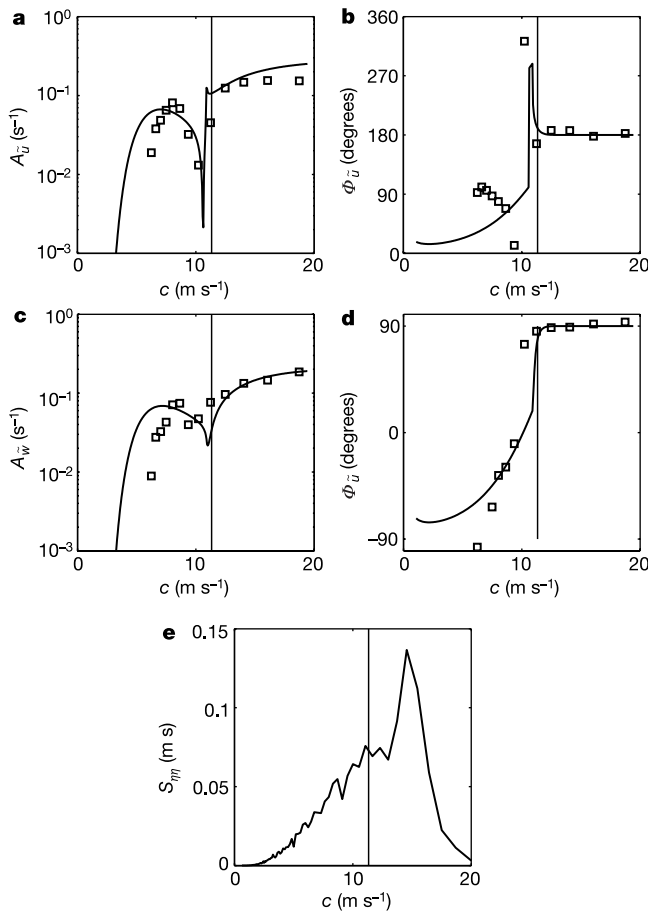


Figure 3 Theoretical and experimental transfer functions for the wave-coherent velocities $T_{\tilde{u}}(c) = A_{\tilde{u}}e^{i\Phi_{\tilde{u}}}$ and $T_{\tilde{w}}(c) = A_{\tilde{w}}e^{i\Phi_{\tilde{w}}}$ and a surface elevation power spectrum.

a, b, Amplitude $A_{\tilde{u}}$ (**a**) and phase shift $\Phi_{\tilde{u}}$ (**b**) for the along-wind wave-induced velocity fluctuations $\tilde{u}$. **c, d**, Amplitude $A_{\tilde{w}}$ (**c**) and phase shift $\Phi_{\tilde{w}}$ (**d**) for the vertical wave-induced velocity fluctuations $\tilde{w}$. **(e)**, Surface wave power spectrum $S_{\eta\eta}$ versus the wave phase speed c . The amplitudes in **a** and **c** are normalized by the surface wave amplitude at that phase speed; the phases in **b** and **d** are referenced to the phase of the waves. The solid lines are numerical results from equation (1); the squares are experimental estimates from time series of wind velocity ($z_f = 8.7 \text{ m}$) and wave elevation collected over 90 minutes. The vertical line marks the mean wind speed measured over the same time period, that is, it shows the wave mode for which the instrument is at the mode's critical height.

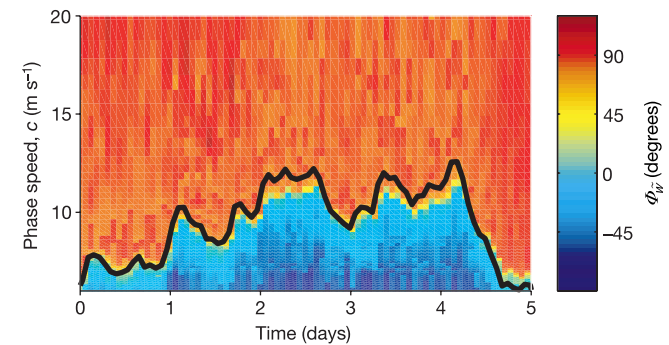


Figure 4 The measured phase of the vertical wave-induced velocity fluctuations $\Phi_{\tilde{w}}$ in degrees throughout a five-day period. The phase is referenced to the phase of the waves. The solid black line shows the mean wind speed $U(z_f)$ versus time, and also marks the wave mode $c = U(z_f)$, for which the instrument is at the mode's critical height. For modes faster than the mean wind the instrument is below their critical heights and for modes slower than the wind the instrument is above their critical heights. The line of the mean wind speed marks the persistent abrupt change in phase, as predicted by the solutions of the Rayleigh equation (1), and in agreement with Figs 1d and 3d. Data are from the anemometer at $z_f = 3.5 \text{ m}$.

Figure 3 shows numerical results and the experimental estimates for the amplitude and phase of the horizontal and vertical wave-coherent wind velocities versus c at fixed z and u_* . For the numerical $\tilde{w}(z, c/u_*)$ at constant z and u_* the abrupt change of its phase is clearly recognizable (Fig. 3d), occurring at $c = U(z_i)$, z_i being the vertical position of the instrument. Indeed, for the wave mode $c = U(z_i)$ the instrument is at the mode's critical height; for all the slower modes $c < U(z_i)$ the instrument is above their critical heights ($z_i > z_c$), and for all faster modes $c > U(z_i)$ the instrument is below their critical heights ($z_i < z_c$). The numerical and the experimental results seem to agree closely and the distinct wave-induced flow features, such as the abrupt change of the phase and the minimum of the amplitude about $c = U(z_i)$, are clearly captured by the experimental data. Nonlinear interaction between wave modes transfers most of the wave energy to the long (fast) waves and leaves little energy in the short (slow) modes, as illustrated by the wave spectrum $S_{\eta\eta}(c)$ in Fig. 3e. Also, as the wave-induced field's vertical decay is scaled by the wavelength, at a given height z_i the shorter the wave the stronger the attenuation of the velocity induced by the wave. In combination, these two circumstances result in a low signal-to-noise ratio for the short-wave signature in the wind and greater uncertainty for the estimated amplitude and phase of the velocity induced by these short (slow) waves. Therefore, numerics and experiment show closer agreement for long (fast) waves and greater divergence for short (slow) waves (Fig. 3a–d).

Throughout the experiment the wave-induced flow maintains the critical layer pattern. That pattern is clearly observed in the phase of the vertical wave-induced velocity fluctuations (Fig. 4), which exhibits a distinct and persistent change along the mean wind line $c = U(z_i)$. For fast ($c > U(z_i)$) waves the phase remains close to 90°, while for waves slower than the mean wind ($c < U(z_i)$) it rapidly decreases, as predicted by the numerical results in Figs 1d and 3d.

Thus we identified the wave-induced flow from field measurements and showed that its configuration is closely consistent with the predictions of the critical-layer theory³ for the range $16 < c/u_* < 40$. Such a result confirms that for the resolved wave scales the critical-layer mechanism of wind–wave coupling is clearly active over the open ocean. Models of climate as well as weather and wave forecasting¹⁹ rely on current knowledge about air–sea fluxes as boundary conditions on the ocean–atmosphere interface. Data from measurements have been used in extensive efforts to parameterize the air–sea momentum flux by expressing the ocean surface drag coefficient $C_D = u_*^2/U^2$ through a single variable. There, the experimental points have systematically failed to collapse onto a particular curve^{10–12} and the observed scatter has not been reduced by accumulating more statistics and by improving the quality of measurements. The evidence presented here in support of the critical layer mechanism³ indicates (i) that the total wave-induced momentum flux $\tilde{\tau} \equiv -\rho\langle\tilde{u}\tilde{w}\rangle = \tilde{E}/c$ is controlled by the wave spectrum and not by a single representative parameter and (ii) that $\tilde{\tau}$ has a considerable vertical variability. These two circumstances are probably major contributors to the observed scatter of drag coefficient estimates. □

Received 24 May; accepted 17 December 2002; doi:10.1038/nature01382.

- Thomson, W. (Lord Kelvin) Hydrokinetic solutions and observations. *Phil. Mag.* **42**(4), 362–377 (1871); *Math Phys. Pap.* **4**, 80.
- Jeffreys, H. On formation of waves by wind. *Proc. R. Soc. Lond. A* **107**, 189–206 (1924).
- Miles, J. W. On the generation of surface waves by shear flows. *J. Fluid Mech.* **3**, 185–204 (1957).
- Belcher, S. & Hunt, J. C. R. Turbulent shear flow over slowly moving waves. *J. Fluid Mech.* **251**, 109–148 (1993).
- Meirink, J. F. & Makin, V. Modelling low-Reynolds-number effects in the turbulent air flow over water waves. *J. Fluid Mech.* **415**, 155–174 (2000).
- Li, P. Y., Xu, D. & Taylor, P. A. Second order turbulence closure modeling for air flow over water waves. *Boundary-Layer Meteorol.* **95**, 397–425 (2000).
- Sullivan, P., McWilliams, J. & Moeng, C.-H. Simulation of turbulent flow over idealized water waves. *J. Fluid Mech.* **404**, 47–85 (2000).
- Hristov, T., Friehe, C. & Miller, S. Wave-coherent fields in air flow over ocean waves: Identification of cooperative behavior buried in turbulence. *Phys. Rev. Lett.* **81**, 5245–5248 (1998).

- Phillips, O. M. On the generation of waves by turbulent wind. *J. Fluid Mech.* **2**, 417–445 (1957).
- Donelan, M. *Air–Sea Interaction* Vol. 9 (eds LeMehaute, B. & Hanes, D. M.) 239–292 (John Wiley & Sons, New York, 1990).
- Komen, G., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S. & Janssen, P. *Dynamics and Modelling of Ocean Waves* (Cambridge Univ. Press, Cambridge, 1994).
- Csanady, G. T. *Air–Sea Interaction: Laws and Mechanisms* (Cambridge Univ. Press, Cambridge, 2001).
- Longuet-Higgins, M. S. Action of variable stress at the surface of water waves. *Phys. Fluids* **12**, 737–740 (1969).
- Davis, R. E. On the turbulent flow over a wavy boundary. *J. Fluid Mech.* **42**, 721–731 (1970).
- Landau, L. D. & Lifshitz, E. M. *Fluid Mechanics* 172–176 (Pergamon, Oxford, 1986).
- Lin, C. *The Theory of Hydrodynamic Stability* (Cambridge Univ. Press, Cambridge, 1955).
- Phillips, O. M. *The Dynamics of the Upper Ocean* 108 (Cambridge Univ. Press, Cambridge, 1977).
- Hasse, L. & Dobson, F. *Introductory Physics of the Atmosphere and Ocean* 101 (D. Reidel, Boston, 1986).
- Janssen, P. et al. *Impact and Feedback of Ocean Waves on the Atmosphere* Vol. 1 of *Advances in Fluid Mechanics* Ch. 5 155–198 (WIT, Ashurst, 2002).

Acknowledgements J. Edson's help and the expert assistance of FLIP's captain Tom Golfinos and crew made it possible to carry out the experiment. In discussions, comments or criticisms came from O. Phillips, M. Donelan, J. McWilliams, P. Sullivan, J.C.R. Hunt, P. Taylor, M. McIntyre, S. Belcher, V. Makin and V. Kudryavtsev. ONR (Marine Meteorology) has provided the funding for this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.S.H. (e-mail: Tihomir.Hristov@jhu.edu).

Iron–silica interaction at extreme conditions and the electrically conducting layer at the base of Earth's mantle

L. Dubrovinsky*, N. Dubrovinskaia*, F. Langenhorst*, D. Dobson*, D. Rubie†, C. Geßmann†, I. A. Abrikosov‡, B. Johansson‡§, V. I. Baykov§, L. Vitos☆, T. Le Bihan¶, W. A. Crichton¶, V. Dmitriev# & H.-P. Weber#

* Bayerisches Geoinstitut, Universität Bayreuth, D-95440 Bayreuth, Germany

† Max-Planck-Institut für Chemie, Joh.-Joachim-Becher-Weg 27, D-55128 Mainz, Germany

‡ Condensed Matter Theory Group, Department of Physics, Uppsala University, S-751 21 Uppsala, Sweden

§ Applied Materials Physics, Department of Materials Science and Engineering, Royal Institute of Technology, Brinellvägen 23, SE-100 44, Stockholm, Sweden

¶ Theoretical Physics Department, Moscow State Institute of Steel and Alloys, Leninskii pr., 4, 117936 Moscow, Russia

¶ European Synchrotron Radiation Facility,

Swiss-Norwegian Beam Lines at ESRF, F-38043 Grenoble, France

☆ Research Institute for Solid State Physics and Optics, H-1525 Budapest, P.O. Box 49, Hungary

The boundary between the Earth's metallic core and its silicate mantle is characterized by strong lateral heterogeneity and sharp changes in density, seismic wave velocities, electrical conductivity and chemical composition^{1–7}. To investigate the composition and properties of the lowermost mantle, an understanding of the chemical reactions that take place between liquid iron and the complex Mg–Fe–Si–Al-oxides of the Earth's lower mantle is first required^{8–15}. Here we present a study of the interaction between iron and silica (SiO₂) in electrically and laser-heated diamond anvil cells. In a multianvil apparatus at pressures up to 140 GPa and temperatures over 3,800 K we simulate conditions down to the core–mantle boundary. At high temperature and pressures below 40 GPa, iron and silica react to form iron oxide and an iron–silicon alloy, with up to 5 wt% silicon. At pressures

of 85–140 GPa, however, iron and SiO_2 do not react and iron–silicon alloys dissociate into almost pure iron and a CsCl-structured (B2) FeSi compound. Our experiments suggest that a metallic silicon-rich B2 phase, produced at the core–mantle boundary (owing to reactions between iron and silicate^{2,9,10,13}), could accumulate at the boundary between the mantle and core and explain the anomalously high electrical conductivity of this region⁶.

Although iron (with ~5 wt% Ni) is a dominant component of the Earth's core, Fe–Ni alloy is too dense by ~10% for the outer liquid core and by 2–5% for solid inner core to satisfy the observed density along any reasonable geotherm¹⁶. Thus, on the basis of cosmochemistry, it has been proposed that the core also contains one or more light elements, such as H, C, O, S, and/or Si¹⁷. It is likely that such light elements were dissolved into the liquid metal during core formation in a magma ocean⁸ during the early history of the Earth. At the probable temperature of the magma ocean (~2,800 K), 2–6 wt% Si can dissolve in liquid Fe at 25 GPa and an appropriate oxygen fugacity⁸. On the basis of a simple thermodynamic model, it was proposed that the solubility of Si at core conditions is close to zero⁸. Thus, after core formation, Si should be expelled from liquid Fe as the metallic core evolves towards chemical equilibrium. In particular, it was demonstrated¹⁸ that

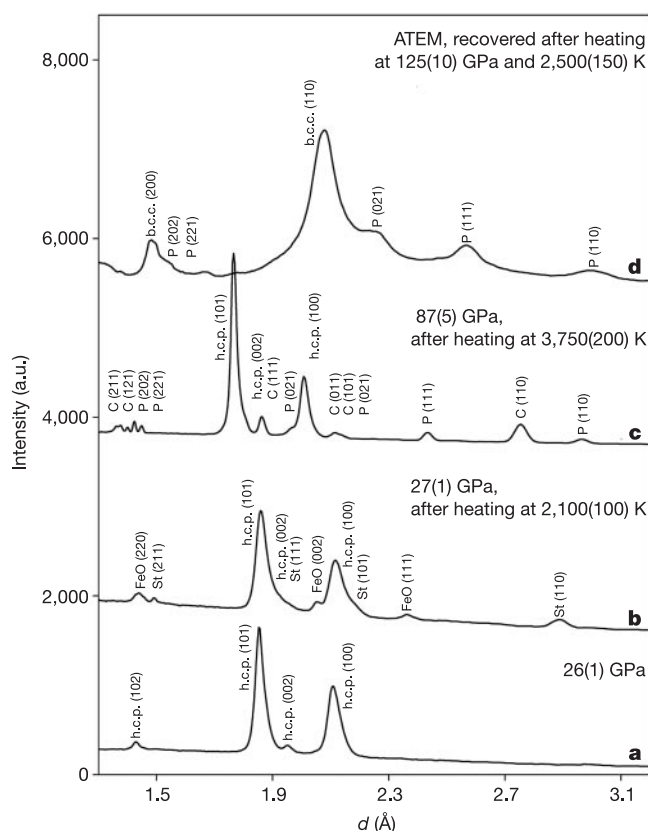


Figure 1 Representative spectra obtained in experiments with iron and amorphous silica as starting materials. **a–c**, Angle-dispersive X-ray and **d**, electron diffraction patterns. **a**, At 26(1) GPa (where 26(1) represents 26 ± 1) before heating of the sample, only diffraction lines of ϵ -Fe (marked as h.c.p.) are present. **b**, After heating at 2,100(100) K silica crystallized in stishovite (St) and new lines of wüstite (FeO) appeared. **c**, Heating of iron and amorphous SiO_2 mixture at 85(5) GPa and 2,400(150) K resulted in crystallization of silica phases only (CaCl₂- and α -PbO₂-structures, marked (C) and (P), correspondingly), but wüstite was absent. **d**, Selected area electron diffraction spectrum of the material recovered after heating at 125(10) GPa and 2,500(150) K shows only the presence of α -PbO₂-type (P) silica phase and iron (b.c.c.). ATEM, analytical transmission electron microscopy; h.c.p., hexagonal close packed; b.c.c., body-centred cubic.

iron can react with MgSiO_3 -perovskite at high pressures and temperatures. The chemical nature of this process is reduction of silicon by more electropositive iron with formation of iron oxide and iron silicide¹⁸.

Two important observations regarding high-pressure phase relations in this Fe–Si system have recently been reported: the synthesis of iron silicide with the CsCl (B2) structure at 24 GPa and high temperatures¹⁹, and the stabilization of the body-centred cubic (b.c.c.) structure in iron-rich Fe–Si alloys at pressures over 80 GPa and temperatures up to 2,400 K (ref. 17). In the light of these recent discoveries, we have further examined chemical reactions between Fe and SiO_2 at conditions of the Earth's mantle down to the core–mantle boundary (CMB).

We conducted three experiments in a multi-anvil apparatus (see Supplementary Information) on a mixture of Fe and SiO_2 , contained in MgO capsules, at 22 GPa and 2,473 K. Electron microprobe analysis shows that the products of all experiments are similar and consist of Fe–Si alloy with compositions ranging from 3.5 to 4.9 wt% Si, magnesiowüstite (a reaction product of the MgO capsule material), and $(\text{Mg,Fe})_2\text{SiO}_4$ (found at the capsule–sample

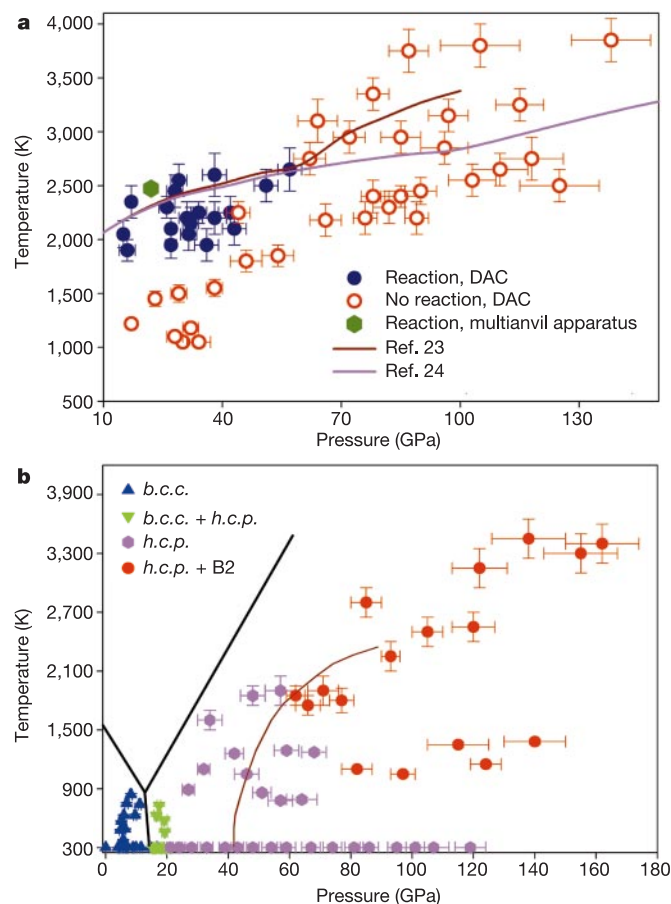


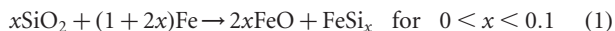
Figure 2 Experimental results on the chemical interaction of iron and silica and behaviour of Fe–Si alloys at high pressures and temperatures. **a**, Filled blue dots show pressure and temperature conditions at which the reaction between iron and silica was observed, and open red circles show conditions at which the reaction does not occur. The green hexagon corresponds to observation of the reaction in multi-anvil experiments. Solid lines show melting curve of iron in pink²³ and dark red²⁴. **b**, Phases observed in electrically-heated (below 1,500 K) and laser-heated (above 1,500 K) diamond anvil cell (DAC) experiments with iron–silicon alloys: 9.6(1), 5.1(1), 4.3(4) wt% Si. We found that all alloys behave similarly and symbols show only phase relations. Continuous black lines show phase relations in pure iron. The red curve shows a phase boundary between h.c.p. and h.c.p. + b.c.c. regions in a Fe–7.9 wt% Si alloy¹⁷.

interface). A sample reacted at high pressure and temperature for 90 min contains a very homogeneous 4.3(4) wt% Si (where 4.3(4) represents 4.3 ± 0.4 ; average of 30 data points) metallic portion that is ~ 0.5 mm in diameter. The Fe-Si alloy from this experiment was used in subsequent higher-pressure experiments using the diamond anvil cell (DAC) as described below.

The reaction between iron and silica was also studied *in situ* in the DAC. Iron foil, 5 μ m thick, was sandwiched between two layers of amorphous silica, compressed to high pressures and laser-heated from both sides. Amorphous silica was chosen to increase the reactivity of the oxide. Figure 1 shows representative angle-dispersive X-ray diffraction patterns. At 26(1) GPa and before

heating only diffraction lines of ϵ -Fe are present (Fig. 1a). However, after heating at 2,100(100) K, silica crystallized as stishovite and new lines of wüstite appeared (Fig. 1b). The lattice parameter of iron, quenched after heating at 22–30 GPa, decreased from 2.8624(8) Å to 2.860–2.858 Å, which corresponds to a Fe-Si alloy with 3.6–4.5 wt% Si. However, heating of iron and amorphous SiO₂ (or re-heating of the Fe-Si alloy and stishovite) at 87(5) GPa and 3,750(200) K (for example well above iron melting temperature at corresponding pressure), as shown in Fig. 1c, resulted only in the crystallization of silica phases (CaCl₂- and α -PbO₂-structured²⁰). No trace of wüstite was observed in the quenched sample at high pressure and the lattice parameter of iron (2.8629(9) Å) after decompression was the same as in the starting material. Analytical transmission electron microscopy (ATEM) (see Supplementary Information) with nanometre-scale resolution of the sample after heating at 125(10) GPa and 2,500(150) K shows only the presence of a silica phase and iron (Fig. 1d). Within the analytical detection limit (0.2 wt% Si) there is no silicon in the metallic portion of the sample. In other words, we did not observe a reaction between iron and silicon at pressures of 87 and 125 GPa.

Figure 2 summarizes the results of our experiments on the Fe-SiO₂ reaction. At pressures between 15 GPa and 40 GPa we detected reduction of silicon by iron:



At higher pressures, above 80 GPa, we did not detect any reaction. The reason that iron and silica do not react at high pressure could be the decreasing solubility of Si in ϵ -Fe (hexagonal close-packed (h.c.p.) structure) with increasing pressure²¹. Indeed, *ab initio* calculations (see Supplementary Information) predict that with increasing pressure the amount of Si that can be accommodated by h.c.p.-Fe substantially decreases and the alloy should dissociate to a mixture of silicon-poor h.c.p. Fe and the Si-rich B2 structured phase (see Supplementary Information).

Note that the theoretical calculations were done at $T = 0$ K, whereas at high temperature entropic contribution may be important²². Therefore, theoretical predictions were tested through a series of high pressure and temperature DAC experiments on Fe-Si alloys containing 9.6(1), 5.1(1) and 4.3(4) wt% Si (later produced by equilibrating Fe with SiO₂ at 22 GPa and 2,473 K in multianvil apparatus, as described above). We found that all alloys behave similarly and Fig. 3 shows representative X-ray patterns obtained *in situ* at high pressure and temperature using the 5.1 wt% Si alloy as the starting material. At 300 K and pressures below 14–18 GPa all studied alloys have the b.c.c. structure (Fig. 3a). At higher pressures the alloys start to transform to the h.c.p. phase and above 20–22 GPa the transformation is completed (Fig. 3b). No further phase transformations were observed on compression to over 100 GPa (Fig. 3c) at ambient temperature, but heating promoted a transformation. Figure 3d (see also left inset in Fig. 3) shows an example of a diffraction pattern collected after 5 h of external electrical heating of the Fe-5.1 wt% Si alloy at 140(10) GPa and 1,380(25) K. Two new lines at 2.494 Å and 1.762 Å belong to the B2 structured alloy ($a = 2.493(1)$ Å) as predicted by *ab initio* calculations (see above). When the same sample was heated for 8 h at 30(2) GPa and 1,100(10) K, the extra lines disappeared, indicating that the reaction was reversed and only the h.c.p. phase remained (Fig. 3e).

One of the samples of Fe-10 wt% Si alloy was compressed to 105(5) GPa in a NaCl pressure medium, laser-heated from both sides at 2,500(100) K. After decompression, the NaCl was carefully dissolved in water. The diffraction pattern of the recovered foil is shown in Fig. 3f (see also right inset in Fig. 3). The extra reflections at 1.969 Å and 2.781 Å correspond to the (110) and (100) peaks of quenched B2 FeSi phase. The lattice parameter of this phase (2.783(2) Å) is slightly lower than the lattice parameter of B2-structured FeSi (2.7917(1) Å) synthesized¹⁹ in multianvil experiments,

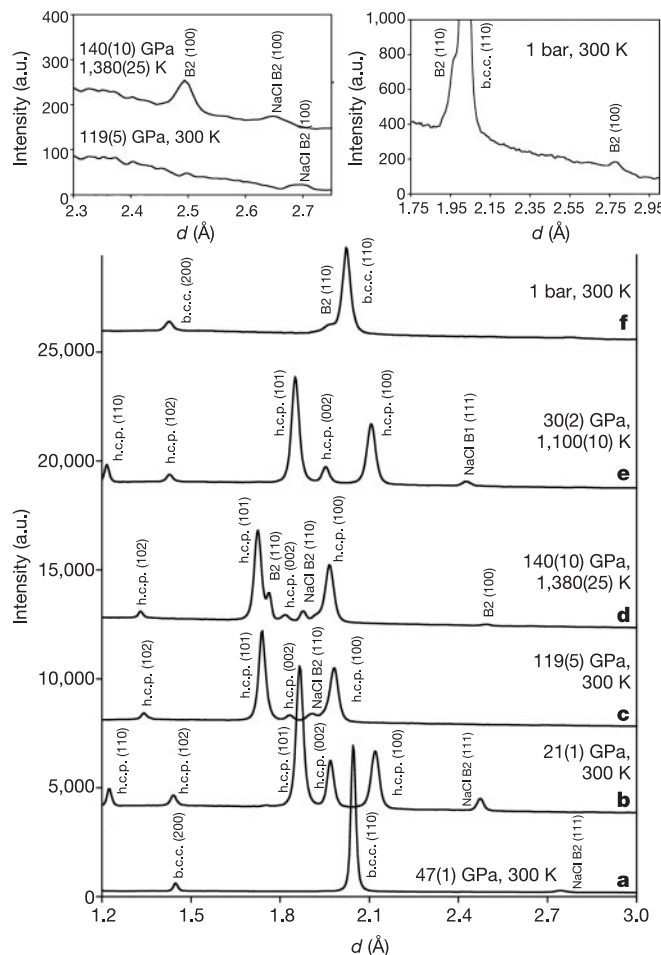


Figure 3 Examples of X-ray diffraction patterns collected in experiments with Fe-5.1 wt% Si. **a**, At 47(1) GPa the alloy has b.c.c. structure. **b**, At higher pressure (above 14 GPa) it starts to transform to h.c.p. phase and at 21(1) GPa the transformation is completed **c**. No phase transformations were observed on compression to over 119(5) GPa at room temperature **d**. However, after five hours of external electrical heating at 140(10) GPa and 1,380(25) K (see also left inset) two new lines appeared at 2.494 Å and 1.762 Å belonging to the B2 structured alloy ($a = 2.493(1)$ Å). **e**, Then, the same sample was heated for 8 h at 30(2) GPa and 1,100(10) K and the extra lines disappeared owing to back reaction of iron alloys with different concentration and only h.c.p. phase remained. NaCl was used as pressure medium (reflections marked as NaCl B1 below 30 GPa and NaCl B2 at higher pressures). Spectrum **f** and the right inset show an X-ray diffraction pattern of the sample recovered after laser-heating of Fe-10 wt% Si alloy at 105(5) GPa and 2,500(100) K (NaCl was dissolved in water). The reflections at 1.969 Å and 2.781 Å correspond to the (110) and (100) peaks of the quenched B2 FeSi phase. The lattice parameter of this phase ($d = 2.783(2)$ Å) is just slightly lower than the lattice parameter of B2-structured FeSi ($d = 2.7917(1)$ Å) synthesized¹⁹ in multianvil experiments, indicating that the composition of the B2 phase obtained in our experiments is close to 1:1 Fe:Si mole ratio.

indicating that composition of the B2 phase obtained in our experiments is close to 1:1 Fe:Si molar ratio.

As mentioned above, the Fe-4.3(4) wt% Si alloy, obtained by reacting SiO₂ and Fe at 22 GPa and 2,473 K in a multianvil apparatus, was homogeneous. However, ATEM studies of the material recovered after treatment of this alloy in a laser-heated DAC at 93(3) GPa and 2,100–2,400 K reveal a variation of silicon concentrations in different parts of the sample from ~0 to 6.6 wt%. These observations provide direct proof that the Fe–Si alloy, equilibrated with SiO₂ at relatively low pressures (20–30 GPa), dissociates into Si-poor and Si-rich phases at much higher pressures (about 100 GPa).

Our results on the behaviour of the Fe–Si alloys with different Si concentrations are summarized in Fig. 2b. We found that at pressures above 60 GPa and high temperatures the alloys dissociate into a mixture of h.c.p.-structured Si-poor and B2-structured Si-rich phases. Lin *et al.*¹⁷ also reported dissociation of Fe–Si alloys with a Si content of more than 4 wt% on the mixture of h.c.p. and b.c.c. phases. The difference between b.c.c.- and B2-structured alloys is solely the ordering of Fe/Si atoms among the positions in the body-centred lattice—fully or partially ordered alloys are B2, and completely disordered alloys are b.c.c. Our diffraction data clearly show the (100) reflection of the B2 structure (Fig. 3d and insets), which is expected to be low in intensity (calculated intensity is 14% for a fully ordered stoichiometric FeSi compound), and could be rather difficult to detect in high pressure and temperature experiments.

The reaction between iron and silicate^{2,9,10,13} could be a source of the iron–silicon alloy at the CMB. The density of B2 FeSi (~9.0 g cm⁻³) is significantly higher than the density of the mantle (~5.6 g cm⁻³) above the CMB and lower than the density of the core (~10.0 g cm⁻³) immediately below the CMB. This means that the silicon-rich alloy would accumulate at the boundary between mantle and core. *Ab initio* simulations (see Supplementary Information) and measurements on B2 FeSi recovered from multianvil experiments¹⁹ show that this compound is an electric conductor (with the electrical conductivity of $6(1) \times 10^5$ S m⁻¹ measured at ambient conditions). Thus, anomalously high electrical conductivity at the base of Earth's mantle⁶ could be associated with the presence of B2 FeSi.

We suggest that if silicon is present in the inner core, it should form a B2 FeSi phase (assuming that there are no further phase transitions in this material above 300 GPa and high temperatures). This implies that the Earth's inner core can be compositionally heterogeneous and contains at least two phases—the Fe-rich phase with hexagonal symmetry and the Fe–Si phase with cubic symmetry. The elastic and rheological properties of these two phases are expected to be quite different and could influence the understanding of the observed inner core anisotropy and heterogeneity^{5,6,11,16,17}. □

Received 28 May 2002; accepted 8 January 2003; doi:10.1038/nature01422.

- Gurnis, M., Wyssession, M. E. & Knittle, E. *The Core-Mantle Boundary Region* (ed. Buffett, B. A.) (AGU, Washington DC, 1998).
- Jeanloz, R. & Williams, Q. in *Ultra-high-Pressure Mineralogy: Physics and Chemistry of the Earth's Deep Interior* (ed. Hemley, R. J.) *Rev. Mineral.* **37**, 241–259 (1998).
- Buffett, B. A., Garnero, E. J. & Jeanloz, R. Sediments at the top of Earth's core. *Science* **290**, 1338–1342 (2000).
- Li, J. & Agee, C. B. Geochemistry of mantle–core differentiation at high pressure. *Nature* **381**, 686–689 (1996).
- Vidale, J. E. & Earle, P. S. Fine-scale heterogeneity in the Earth's inner core. *Nature* **404**, 273–275 (2001).
- Buffett, B. A. Constraints on magnetic energy and mantle conductivity from the forced nutations of the Earth. *J. Geophys. Res.* **97**, 19581–19597 (1992).
- Carnero, E. J. & Jeanloz, R. Fuzzy patches on the Earth's core-mantle boundary? *Geophys. Res. Lett.* **27**, 2777–2780 (2000).
- Gessmann, C. K., Wood, B. J., Rubie, D. C. & Kilburn, M. R. Solubility of silicon in liquid metal at high pressure: implications for the composition of the Earth's core. *Earth Planet. Sci. Lett.* **184**, 367–376 (2001).
- Knittle, E. & Jeanloz, R. Simulating the core-mantle boundary: an experimental study of high-pressure reactions between silicates and liquid iron. *Geophys. Res. Lett.* **16**, 609–612 (1989).
- Poirier, J. P., Malavergne, V. & Le Mouél, J. L. in *The Core-Mantle Boundary Region* (eds Gurnis, M., Wyssession, M. E., Knittle, E. & Buffett, B. A.) 131–137 (AGU, Washington DC, 1998).

- Steinle-Neumann, G., Stixrude, L., Cohen, R. E. & Gulseren, O. Elasticity of iron at the temperature of the Earth's inner core. *Nature* **413**, 57–60 (2001).
- Goarant, F., Guyot, F., Peyronneau, J. & Poirier, J. P. High-pressure and high-temperature reactions between silicates and liquid iron alloys in the diamond anvil cell, studied by analytical electron microscopy. *J. Geophys. Res.* **97**, 4477–4487 (1992).
- Song, X. & Ahrens, T. J. Pressure-temperature range of reactions between liquid iron in the outer core and mantle silicates. *Geophys. Res. Lett.* **21**, 153–156 (1994).
- Tschauner, O., Zerr, A., Specht, S., Rocholl, A., Boehler, R. & Palme, H. Partitioning of nickel and cobalt between silicate perovskite and metal at pressures up to 80 GPa. *Nature* **398**, 604–607 (1999).
- Dubrovinsky, L. *et al.* Chemical interaction of iron and corundum as a source of heterogeneity at the core-mantle boundary. *Nature* **412**, 527–529 (2001).
- Dubrovinsky, L. S., Saxena, S. K., Tutti, F. & Le Bihan, T. X-ray study of thermal expansion and phase transition of iron at megabar pressure. *Phys. Rev. Lett.* **84**, 1720–1723 (2000).
- Lin, J.-F., Heinz, D. L., Campbell, A. J., Devine, J. M. & Shen, G. Iron-silicon alloy in Earth's core? *Science* **295**, 313–315 (2002).
- Knittle, E. & Jeanloz, R. Earth's core-mantle boundary: Results of experiments at high pressures and high temperatures. *Science* **251**, 1438–1443 (1991).
- Dobson, D. P., Vocadlo, L. & Wood, I. G. A new high-pressure phase of FeSi. *Am. Mineral.* (in the press).
- Prokopenko, V. B., Dubrovinsky, L. S., Dmitriev, V. & Weber, H.-P. Raman spectroscopy and X-ray diffraction *in situ* characterization of phase transitions in cristobalite under high pressure. *J. Alloys Compounds* **327**, 87–95 (2001).
- Guyot, F. *et al.* P-V-T measurements of iron silicide (e-FeSi)—Implications for silicate-iron interactions in the early Earth. *Eur. J. Mineral.* **9**, 277–285 (1997).
- Alfe, D., Price, G. D. & Gillan, M. J. Thermodynamic stability of Fe/O solid solution at inner-core conditions. *Geophys. Res. Lett.* **27**, 2417–2420 (2000).
- Boehler, R. Temperatures in the Earth's core from melting-point measurements of iron at high static pressures. *Nature* **363**, 534–536 (1993).
- Shen, G., Mao, H. K., Hemley, R. J. & Rievers, M. L. Melting and crystal structure of iron at high pressure. *Geophys. Res. Lett.* **25**, 373–376 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank G. Herrmannsdörfer, H. Schulze, R. Eger and C. Kamella, for technical assistance. Discussion with G. D. Price allowed us to improve the manuscript. Financial support of Deutsche Forschungsgemeinschaft, the Swedish Research Council (VR), the Swedish Foundation for Strategic Research (SSF), and the EU 'Access to Research Infrastructures' Programme is acknowledged.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.D. (e-mail: Leonid.Dubrovinsky@uni-bayreuth.de).

A Middle Miocene hominoid from Thailand and orangutan origins

Yaowalak Chaimanee*, Dominique Jolly†, Mouloud Benammi‡, Paul Tafforeau§, Danielle Duizer†, Issam Moussa§ & Jean-Jacques Jaeger§

* Paleontology Section, Geological Survey Division, Department of Mineral Resources, Bangkok -10400, Thailand

† Palynologie et Paléoenvironnements, I.S.E.M., cc 061, Place Eugene Bataillon, 34095-Montpellier, France

‡ Paleomagnetismo, Instituto de Geofísica, Universidad Nacional Autónoma de México, 04510 México DF, México

§ Paléontologie, I.S.E.M., cc 064, Université Montpellier II, Place Eugene Bataillon, 34095-Montpellier, France

The origin of orangutans has long been debated. *Sivapithecus* is considered to be the closest ancestor of orangutans because of its facial–palatal similarities¹, but its dental characteristics² and postcranial skeleton^{2,3} do not confirm this phylogenetic position. Here we report a new Middle Miocene hominoid, cf. *Lufengpithecus chiangmuanensis* n. sp. from northern Thailand. Its dental morphology relates it to the *Pongo* clade, which includes *Lufengpithecus*^{4,5}, *Sivapithecus*², *Gigantopithecus*⁶, *Ankarapithecus*⁷ and possibly *Griphopithecus*⁸. Our new species

displays striking dental resemblances with living orangutans and appears as a more likely candidate to represent an ancestor of this ape. In addition, it originates from the geographic area of Pleistocene orangutans. But surprisingly, the associated flora shows strong African affinities, demonstrating the existence of a temporary floral and faunal dispersal corridor between south-east Asia and Africa during the Middle Miocene, which may have played a critical role in hominoid dispersion.

The new hominoid remains were recovered from the middle and upper lignite seams of Ban Sa locality, Chiang Muan basin, northern Thailand. The sediments consist mainly of mudstones and sandstones intercalated with palaeosols and several lignite beds. Associated large mammals are similar to those of the Chinji Formation of Pakistan, dated between 14 and 10.8 million years (Myr)⁹. Palaeomagnetic data indicate two normal events sandwiching a reverse polarity zone, the hominoid originating from the upper normal zone (Fig. 1). Our age estimation, between 13.5 and 10 Myr, is congruent with the absence of C4 plants, which appear around 10 Myr in the Siwaliks¹⁰ and become abundant about 8–7 Myr¹¹. Enamel of several ungulates shows ¹²C/¹³C isotopic ratios ($\delta^{13}\text{C}$ values) of –13.4 to –11.8 ‰ (Pee Dee belemnite standard), demonstrating a diet exclusively based on C3 plants¹⁰. Palaeosols indicate some climatic seasonality, which is confirmed by palynology from the lignites. The vegetation corresponded to a mosaic of tropical freshwater swamps, with a *Syzygium*-dominated lowland forest (Fig. 2), similar to an extant African habitat from the White Nile headwaters¹², but differing from the temperate Late Miocene Lufeng flora of southern China¹³. This resemblance with African flora testifies to the existence, in the Middle Miocene, of a dispersal corridor between southeast Asia and Africa, the duration and location of which have yet to be precisely determined.

Order Primates Linnaeus, 1758
Suborder Anthropoidea Mivart, 1864
Superfamily Hominoidea Gray, 1825
Family Hominidae Gray, 1825
Subfamily Ponginae Elliot, 1913
cf. *Lufengpithecus chiangmuanensis* sp. nov.

Etymology. The species is named according to the type locality.

Holotype. Right lower canine fragment (TF 6171-1), left P₃ fragment (TF 6171-2), right P₄ (TF 6171-3), right M₂ (TF 6171-4), left M₂₋₃ (TF 6171-5, 6171-6) (Fig. 3 c–f, h) and right P₃ fragment (TF 6171-7) (Department of Mineral Resources, Bangkok).

Referred materials. Right I¹ (TF 6168), right M² (TF 6169), left dP₄ (TF 6170) (Fig. 3 a, b, g). Left upper canine (TF 6174), right I² (TF 6173), left I₁ (TF 6178), right P³ (TF 6175), left M²⁻³ (TF 6176, 6177), left P₄ (TF 6179) and right M₂ (TF 6180) (Fig. 3 i–o) and right M₃ fragment (TF 6172).

Horizon and locality. Middle and upper lignite seams of Chiang Muan Basin, Ban Sa locality, northern Thailand, Middle Miocene age.

Diagnosis. Large hominoid with dental size and jugal teeth morphology, wear patterns, crown elevation, enamel thickness and wrinkling, apparent strong sexual dimorphism, similar to *Lufengpithecus lufengensis* Wu 1987, from which it differs by its anterior dentition, greater mesio-distal length and lower elevation of its lower central incisor crowns with more divergent mesial and distal margins, rounded lower canine crown section and less heteromorphic P³, P₃ without metaconid, more slanted P₃ and P₄ buccal cusps and M₃ larger than M₂. Differs from *Sivapithecus* by less linguallly flared upper molar crowns, more nearly square molar occlusal surfaces, higher molar cusps, dentine horns with greater

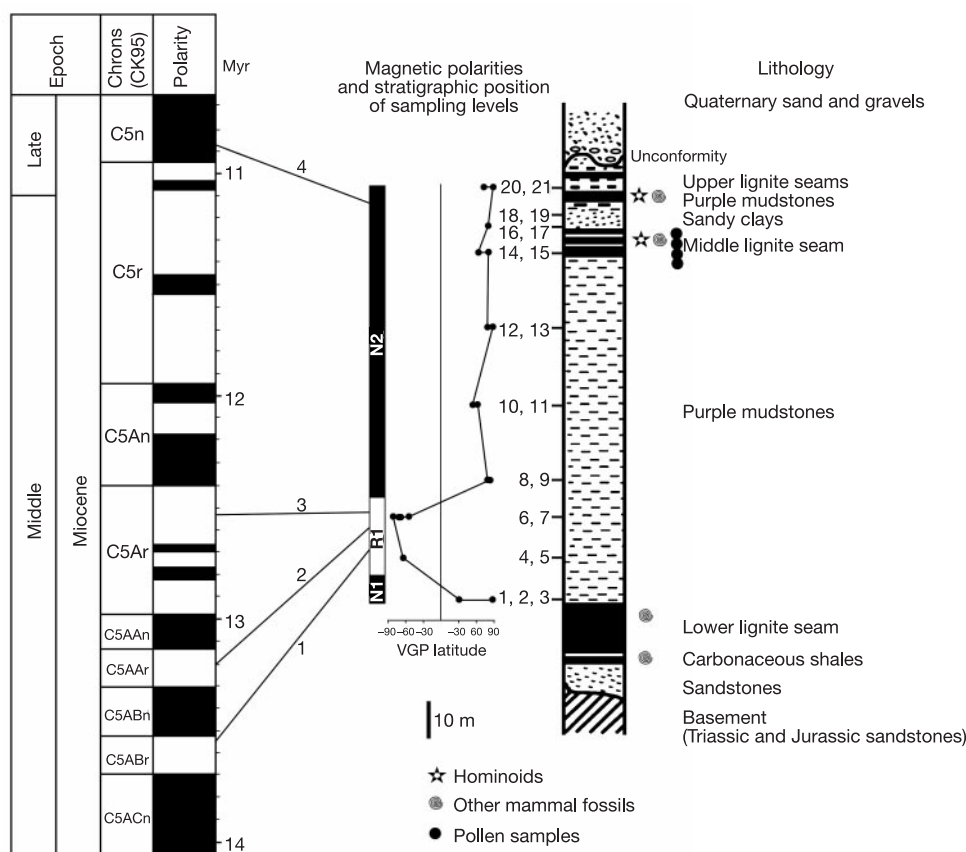


Figure 1 Stratigraphic section of Ban Sa coalmine bearing cf. *Lufengpithecus chiangmuanensis*. Lines (1–4) indicate four possible correlations with the magnetic polarity timescale (see Methods).

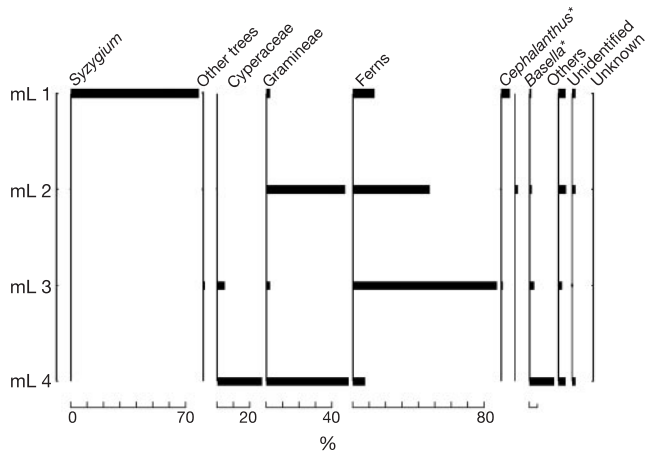


Figure 2 Simplified pollen diagram of four samples collected from the middle lignite seam in Ban Sa section. The pollen spectra (268–1040 identified pollen grains counted in each sample) are characterized by a high taxonomic diversity (30 taxa). *Syzygium* forest dominates at the level that has yielded most hominoid remains (mL1). 'Other trees' includes *Alchornea*, *Celtis*, *Macaranga*, *Muraya*, *Meliaceae*, *Mimosaceae*, *Pometia*, *Rutaceae* and *Moraceae*. 'Others' includes *Rubiaceae*, *Apiaceae*, *Celastraceae*, *Compositae*, *Polygonum*, *Combretaceae*, *Merremia*, *Elitranthe*, *Malpighiaceae*, *Papilionoideae*, *Euphorbiaceae*, *Lasianthus* and *Eleagnaceae*. The presence of *Cephalanthus*, a deciduous wetland shrub, associated with *Celtis*, *Mimosaceae* and some *Combretaceae*, implies a dry-season climate.

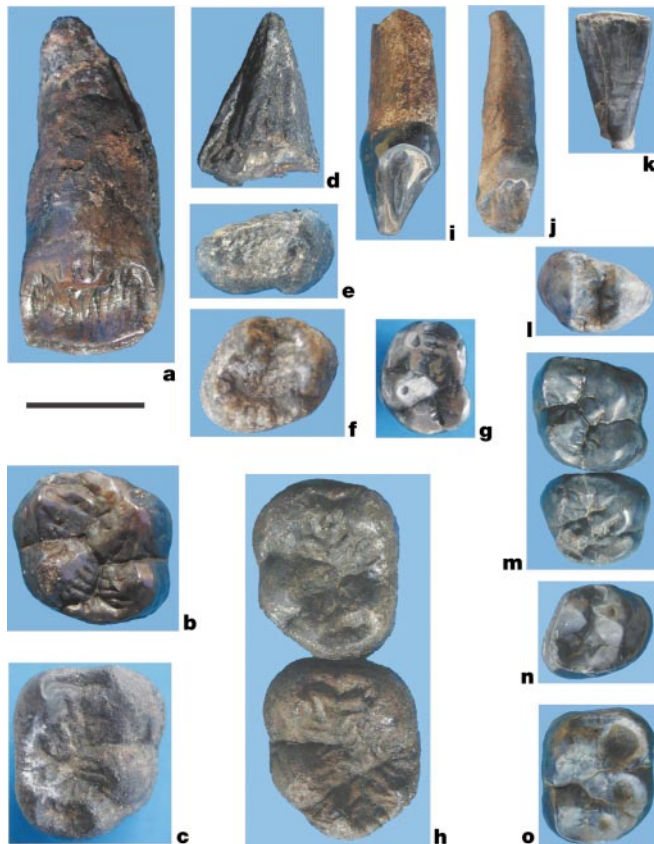


Figure 3 cf. *Lufengpithecus chiangmuanensis* n. sp. **a**, TF 6168, right I¹. **b**, TF 6169, right M². **c**, TF 6171-4, right M². **d**, TF 6171-1, right lower canine. **e**, TF 6171-2, distal part of left P₃ (mirror image). **f**, TF 6171-3, right P₄. **g**, TF 6170, left dP₄. **h**, TF 6171-5, 6171-6, left M₂₋₃. **i**, TF 6174, left upper canine. **j**, TF 6173, right I². **k**, TF 6178, left I₁. **l**, TF 6175, right P³. **m**, TF 6176, 6177, left M²⁻³ (mirror image). **n**, TF 6179, left P₄ (mirror image). **o**, TF 6180, right M₂. Scale bar, 1 cm.

relief and by different morphology of its I¹, I², canine and P³. Differs from *Ankarapithecus* by its larger upper central incisors, smaller P³ and M₃ larger than M₂. Differs from *Griphopithecus* by its larger size and more wrinkled enamel, larger I¹ with different root/crown proportions, less heteromorphic P³ with stronger lingual flaring, molars with more elevated crowns, I₁ with more divergent mesial and distal margins, P₃ without strong beaks and M₃ larger than M₂, without cingulids. Differs from *Pongo* by its less peripheralized and more flared upper lingual and lower buccal molar cusps and greater molar occlusal relief, smaller lower incisors but with similar length/breadth index, less heteromorphic P³ with low and slanted protocone and less densely crenulated enamel. Differs from *Ouranopithecus* by its anterior dentition, less enlarged M₃ and much thinner enamel. Differs from *Dryopithecus* by its thicker enamel, incisors size, absence of cingula, wider lower molars, and smaller P₃.

The holotype teeth belonged to a single individual. Main characters (wrinkled enamel, very large I¹, small and peg-like I², low crowned P³ with reduced heteromorphy, absence of molar cingula) indicate that it belongs to the *Pongo* clade, represented by *Lufengpithecus* (10–3 Myr) (ref. 5), *Sivapithecus* (12.3–8.5 Myr) (ref. 9), *Gigantopithecus* (7–6 Myr), *Ankarapithecus* (9.89–9.59 Myr)⁷ and Pleistocene to extant *Pongo*¹⁴. Two size categories are represented in our samples (Table 1), which, according to the morphological similarities, are interpreted as presumed male and female, as for *L. lufengensis*¹⁵.

I¹ is similar in crown proportions and size to the largest I¹'s of *S. sivalensis* (bucco-lingual breadth/mesio-distal length, 0.77; labial height/mesio-distal length, 0.95) (ref. 16), but its crown, root morphology and root/crown proportions are more similar to *L. lufengensis* (J. Kelley, personal communication). It also resembles *Pongo*, differing by its smaller size, the absence of a strong central lingual pillar and by having fewer vertical lingual ridges. I² is small and peg-shaped, with an asymmetric crown and spiral lingual cingulum, differing from *Pongo* by the absence of a strong central lingual pillar and by stronger asymmetry. Female upper canine is small and low-crowned. It shows weak cervical flare and a well-developed basal lingual cingulum connecting the mesial and distal ridges. P³ has an asymmetric triangular occlusal outline, without cingula and with an extended mesio-buccal corner. Heteromorphy is reduced and the lingual crown surface is strongly flaring. The small-sized M² and M³ share many similarities with those of *L. lufengensis*, differing by their more slanted buccal cusps. M³ is reduced, especially its distal half with a more curving buccal wall than that of M². The anterior fovea is absent and the paracone is

Table 1 Dental measurements (in millimetres) of cf. *Lufengpithecus chiangmuanensis* n. sp. Holotype (TF 6171)

Presumed male		MD	BL	
I ¹ R	(TF 6168)	11.97	8.87	
M ² R	(TF 6169)	12.64	13.48	
		MD	BL (trd.)	(tad.)
dP ₄ L	(TF 6170)	9.92	8.07	
C _R	(TF 6171-1)	10.59	10.54	
P ₃ L	(TF 6171-2)	—	> 8.00	
P ₃ R	(TF 6171-7)	> 12.00	8.55	
P ₄ R	(TF 6171-3)	10.32	10.97	
M ₂ R	(TF 6171-4)	14.38	12.75	11.60
M ₂ L	(TF 6171-5)	13.97	11.76	10.80
M ₃ L	(TF 6171-6)	15.79	12.79	11.80
M ₃ R	(TF 6172)	—	12.70	
Presumed female				
I ² R	(TF 6173)	5.08	4.61	
C _L	(TF 6174)	9.09	8.54	
P ₃ R	(TF 6175)	7.17	9.94	
M ₂ L	(TF 6176)	9.95	10.42	
M ₃ L	(TF 6177)	8.85	10.17	
I ₁ L	(TF 6178)	5.07	5.26	
P ₄ L	(TF 6179)	8.66	7.89	
M ₂ R	(TF 6180)	11.56	9.41	

MD, mesiodistal; BL, buccolingual; trd., trigonid; tad., talonid; R, right; L, left.

more slanted. The male M^2 is larger, strongly wrinkled, but shares the same morphology as the smaller M^2 . I_1 occlusal surface is rather long mesio-distally, its crown is relatively low and the mesial and distal tooth margins are divergent rather than parallel as those of *L. lufengensis*. Only a tiny central rib can be distinguished on its lingual side and there are neither marginal ridges nor lingual cingulum. It shares the same mesio-distal length and bucco-lingual width ratio with *Pongo* (1.05) (Pleistocene *Pongo* (ref. 14): 1.06, $N = 9$, range = 0.89–1.23), which is very different in *L. lufengensis* (ref. 17) (0.55–0.60). I_1 of *Pongo* is larger and has a stronger median lingual rib. The male lower canine is large with the base of its crown missing. Its cross-section is rounded, different from that of *L. lufengensis*. Its lingual wall bears two furrows, mesio-lingual and disto-lingual, bounding a strongly convex median vertical ridge. Canine shape and curvature are different from *L. lufengensis*, being more stout but more similar to that of Pleistocene *Pongo*¹⁴ and sharing a disto-buccal groove with *L. lufengensis*, *Griphopithecus* and *Pongo*. The male P_3 has no metaconid cusp. Its crown is less elevated and bucco-lingually wider with a more linguallly slanted protoconid than that of *L. lufengensis* and there is no cingulid. DP_4 is molariform with a small protoconid, resulting in a narrower mesial end and a strongly curved lingual surface, as in *Griphopithecus*. Its small and narrow trigonid is bent mesially and its paracristid has an arcuate shape as in *Pongo*¹⁸ and *Sivapithecus*². P_4 is similar to that of *L. lufengensis* but displays a longer talonid. It is slightly wider than long, has a linguallly more slanted protoconid wall and its mesial part is wider than the distal. There is no cingulid, as on the lower molars, which are characterized by rounded and elevated cusps and strong enamel wrinkles. Measurements, from micro-computer tomography scans, of relative enamel thickness by Martin's method¹⁹ (Fig. 4) indicate values for male (17.23) and for female (17.80) M_2 that are closer to those obtained for an extant *Pongo pygmaeus* (16.52) and for *Pongo* mean (15.9) (ref. 20) rather than to *Griphopithecus* (19.7) (ref. 20). The morphology of the molars is similar to that of *L. lufengensis*, differing by their more slanted buccal walls and different M_3/M_2 proportions. M_3 size corresponds to that of the largest *L. lufengensis* M_3 s whereas M_2 size is in the middle of the *L. lufengensis* male range¹⁵. M_2 and M_3 have a short buccal cingulid. M_3 entoconid is extremely reduced and its hypoconulid is distally salient and centrally located. The lingual cusps and the hypoconulid of the lower molars have rounded horizontal wear facets, producing isolated and circular dentine pits similar to those of *L. lufengensis* which fuse together in late wear stages, unlike *Sivapithecus*.

To conclude, the new Thai species displays a strong resemblance to *Lufengpithecus*, but differs from it by several dental characters. *L. lufengensis* shows specialized characters in its anterior dentition: lower incisors are narrower mesio-distally, their crowns are higher, their length/width index is much smaller and their lower canine

crowns are more elevated. Its skull^{21–23} and anterior teeth morphology exclude it from being a direct ancestor of orangutans. However, because cf. *L. chiangmuanensis* does not differ more profoundly in comparably preserved parts from *L. lufengensis*, but does appear to be the more closely related to *Pongo*, it might indeed be necessary, especially upon the discovery of more completely preserved specimens, to refer the new species to its own genus. However, pending these discoveries, *Lufengpithecus* would appear as a paraphyletic genus and therefore we propose to refer the new Thai species as cf. *Lufengpithecus chiangmuanensis* n. sp. Despite their more densely crenulated enamel and larger central incisors, Pleistocene and extant orangutans share many characters with cf. *L. chiangmuanensis* but display more derived features. *Sivapithecus* has some postcranial³ and skull characters¹ similar to those of *Pongo* but its dentition is less *Pongo*-like than that of the new Thai species. This new species is older than *L. lufengensis*, is known from the geographic area of Pleistocene *Pongo* and displays characters that make it the closest relative of extant *Pongo*. The dispersal corridor that we have identified, linking tropical Asia with Africa during the Middle Miocene, may also contribute to our understanding of Middle Miocene hominoid phylogeography. □

Methods

Chronology

Twenty-one oriented standard palaeomagnetic drill cores were collected from ten different stratigraphic levels. All samples were measured and analysed at the Laboratorio de Palaeomagnetismo, Universidad Nacional Autonoma de Mexico, on the JR5-A spinner magnetometer. Thermal and alternating fields were used for the demagnetization process. Thermal demagnetization of isothermal remanent magnetization (IRM) shows that magnetite and haematite are the main ferromagnetic minerals. Thermal demagnetization data for each specimen were generated from ten or more temperature steps (or six to nine alternating field steps), and then plotted on orthogonal vector diagrams to identify the characteristic remanent magnetization components (ChRM). The ChRM direction was calculated by using principal component analysis. Demagnetization isolates both normal and reverse polarities with an overall mean direction of $D = 9.8^\circ$, $I = 32.8^\circ$, $N = 14$ and $D = 175.9^\circ$, $I = -37.5^\circ$, $N = 5$, 20 out of 22 samples having been used for this analysis. The obtained data pass the analytical reversal test with 'C' quality classification. The mean direction is close to the expected direction derived from the apparent polar wander path of Eurasia, but the inclination value is lower than expected at the site latitude, consistent with a sedimentary inclination error for an original depositional remanent magnetization. This study highlights the succession of two normal polarities sandwiching a reverse polarity zone, the described hominoids originating from the upper normal zone (N2).

The large mammals are similar to those of the Chinji Formation of Pakistan (14–10.8 Myr). Mammalian fauna consists of stegolophodont mastodon, pig (*Hyotherium* cf. *pilgrimi*) and ruminants (*Eotragus*, *Dorcatherium* and *Siamotragulus*). The pig *H. pilgrimi* shows a temporal distribution in the Siwaliks between 14 and 11.3 Myr (J. Barry, personal communication). The tectonic phase affecting northern Thai intramontane basins is dated about 11.6 Myr (ref. 24). There are four possible correlations, requiring quite variable sedimentation rates through the section. The best fit, taking into account the estimated sedimentation rate²⁵ and the relative length of the reversals, suggests correlation with C5ABn chron corresponding to an age of 13.3–13.5 Myr for the upper normal polarity zone. But the chrons C5ABr and C5ABn are of nearly equal duration, and the sedimentation rate of N1 increases over that of R1. If we correlate R1 with chron C5AAr or chron C5Ar.1r and the corresponding normal chrons with C5AAn and C5An.2n, then it requires even greater differences in sedimentation rates. The last possible correlation is N2 with chron C5n.2n, and in that case, the youngest age for the level bearing cf. *L. chiangmuanensis* would be between 10 and 11 Myr. This age would correspond to the age of early hipparions, which appear in the Siwaliks at 10.7 Myr (ref. 9) and are present in the Late Miocene of Thailand, but absent from Chiang Muan. It appears less likely but still possible according to mammalian biochronology, but would imply a younger age as generally admitted for the tectonic event²⁴. Therefore, the combination of all the available data rather constrains the age of the new Thai hominoid to between 13.5 and 10 Myr.

Received 25 July 2002; accepted 17 January 2003; doi:10.1038/nature01449.



Figure 4 Virtual vertical sections through the mesial cusps of M_2 of presumed male cf. *Lufengpithecus chiangmuanensis* (left) and extant *Pongo pygmaeus* (right) showing comparable relative enamel thickness and dentine penetrance. Note also the stronger flaring of the buccal wall on the Thai fossil. Scale bar, 1 cm.

1. Pilbeam, D. New hominoid skull material from the Miocene of Pakistan. *Nature* **295**, 232–234 (1982).
2. Ward, S. in *Function, Phylogeny, and Fossils: Miocene Hominoid Evolution and Adaptations* (eds Begun, D. R., Ward, C. V. & Rose, M. D.) 269–290 (Plenum, New York, 1997).
3. Pilbeam, D., Rose, M. D., Barry, J. C. & Shah, S. M. I. New *Sivapithecus* humeri from Pakistan and the relationship of *Sivapithecus* and *Pongo*. *Nature* **348**, 237–239 (1990).
4. Wu, R. A revision of the classification of the Lufeng great apes. *Acta Anthropol. Sin.* **6**, 263–271 (1987).
5. Harrison, T., Ji, X. & Su, D. On the systematic status of the Late Neogene hominoids from Yunnan Province, China. *J. Hum. Evol.* **43**, 207–227 (2002).
6. Simons, E. L. & Chopra, S. R. K. *Gigantopithecus* (Pongidae, Hominoidea) A new species from north India. *Postilla* **138**, 1–18 (1969).
7. Kappelman, J. et al. in *Geology and Paleontology of the Sinap Formation* (eds Fortelius, M., Kappelman, J., Bernor, R. & Sen, S.) (Columbia University Press, New York, in press).

8. Alpagut, B., Andrews, P. & Martin, L. New hominoid specimens from the Middle Miocene site at Pasalar, Turkey. *J. Hum. Evol.* **19**, 397–422 (1990).
9. Barry, J. C. *et al.* Faunal and environmental change in the Late Miocene Siwaliks of northern Pakistan. *Paleobiology Memoirs* **3**(28), 1–71 (2002).
10. Morgan, M. E., Kingston, J. D. & Marino, B. D. Carbon isotopic evidence for the emergence of C4 plants in the Neogene from Pakistan and Kenya. *Nature* **367**, 162–165 (1994).
11. Cerling, T. E. *et al.* Global vegetation change through the Miocene/Pliocene boundary. *Nature* **389**, 153–158 (1997).
12. Pajunen, H. *Proc. Symp. Int. Peat Soc., Jamaica, 25 February–1 March 1985* (ed. Heikurainen, L.) 186–197 (Int. Peat Soc., Helsinki, 1985).
13. Badgley, C., Guoqin, Q., Wanyong, C. & Defen, H. Paleoeology of a Miocene, tropical, upland fauna: Lufeng, China. *Nat. Geogr. Res.* **4**, 178–195 (1988).
14. Hooijer, D. A. Prehistoric teeth of man and of the orang-utan from central Sumatra, with notes on the fossil orang-utan from Java and Southern China. *Zool. Med.* **29**, 175–301 (1948).
15. Kelley, J. & Plavcan, J. M. A simulation test of hominoid species number at Lufeng, China: implications for the use of the coefficient of variation in paleotaxonomy. *J. Hum. Evol.* **35**, 577–596 (1998).
16. Kelley, J., Anwar, M., McCollum, M. A. & Ward, S. C. The anterior dentition of *Sivapithecus parvada*, with comments on the phylogenetic significance of incisor heteromorphy in Hominoidea. *J. Hum. Evol.* **28**, 503–517 (1995).
17. Wu, R. & Oxnard, C. E. Ramapithecines from China: evidence from tooth dimensions. *Nature* **306**, 258–260 (1983).
18. Swarts, J. D. in *Orangutan Biology* (ed. Schwartz, J. H.) 263–270 (Oxford University Press, London, 1988).
19. Martin, L. Significance of enamel thickness in hominid evolution. *Nature* **314**, 260–263 (1985).
20. Martin, L. *The Relationships of the Later Miocene Hominoidea*. Thesis, Univ. London (1983).
21. Schwartz, J. H. in *Function, Phylogeny, and Fossils: Miocene Hominoid Evolution and Adaptations* (eds Begun, D. R., Ward, C. V. & Rose, M. D.) 363–388 (Plenum, New York, 1997).
22. Schwartz, J. H. *Lufengpithecus* and its potential relationship to an orang-utan clade. *J. Hum. Evol.* **19**, 591–605 (1990).
23. Harrison, T. & Rook, L. in *Function, Phylogeny, and Fossils: Miocene Hominoid Evolution and Adaptations* (eds Begun, D. R., Ward, C. V. & Rose, M. D.) 327–362 (Plenum, New York, 1997).
24. Remus, D., Webster, M. & Krawkan, K. Rift architecture and sedimentology of the Phetchabun intermontane basin, central Thailand. *J. Southeast Asian Earth Sci.* **8**, 421–432 (1993).
25. Johnson, N. M. & McGee, V. E. Magnetic polarity stratigraphy: stochastic properties of data, sampling problems, and the evolution of interpretations. *J. Geophys. Res.* **88**, 1213–1221 (1983).

Acknowledgements We thank P. Andrews, L. de Bonis, J. Kappelman, J. Kelley and D. Pilbeam for comments, help, discussion and providing comparative materials; J. H. Schwartz and J. Kappelman for improving our manuscript; E. Boller, J. Baruchel and the ID 19 beamline staff of the European Synchrotron Radiation Facility (Grenoble, France) for their help in obtaining microtomographic images; A. Sritulakarn and N. Wongchai for providing facilities in Chiang Muan coal mine; B. Marandat for preparing fossils and making casts; H. Tong for translating Chinese documents; J. Barry, P. Tassy, G. Métais and S. Ducrocq for identifying associated large mammals. This work was supported by the Wenner-Gren and the Fyssen Foundations, the Department of Mineral Resources (Bangkok) and TRF-CNRS Project.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.-J. (e-mail: jaeger@isem.univ-montp2.fr).

Species interactions can explain Taylor's power law for ecological time series

A. M. Kilpatrick & A. R. Ives

Department of Zoology, University of Wisconsin—Madison, Madison, Wisconsin 53706, USA

One of the few generalities in ecology, Taylor's power law^{1–3}, describes the species-specific relationship between the temporal or spatial variance of populations and their mean abundances. For populations experiencing constant per capita environmental variability, the regression of log variance versus log mean abundance gives a line with a slope of 2. Despite this expectation, most species have slopes of less than 2 (refs 2–4), indicating that more abundant populations of a species are relatively less variable than expected on the basis of simple statistical grounds. What causes abundant populations to be less variable has received

considerable attention^{5–12}, but an explanation for the generality of this pattern is still lacking. Here we suggest a novel explanation for the scaling of temporal variability in population abundances. Using stochastic simulation and analytical models, we demonstrate how negative interactions among species in a community can produce slopes of Taylor's power law of less than 2, like those observed in real data sets. This result provides an example in which the population dynamics of single species can be understood only in the context of interactions within an ecological community.

Taylor's power law describes the species-specific scaling relationship between the variance of population abundances and their means, which has been established for more than 400 species in taxa ranging from protists to vertebrates^{2,3}. The ubiquity of Taylor's power-law slopes of between 1 and 2 suggests an underlying fundamental explanation that might give insights into general ecological processes affecting many species. Furthermore, understanding the relationship between a population's variability and its mean abundance is needed to address numerous ecological problems, including conducting population viability analyses in conservation biology¹³, testing diversity–stability hypotheses^{14–16} and sampling for pests in agriculture^{17,18}.

There are two separate relationships between the variance and mean of a species's abundance, one temporal and one spatial³. Of these, the spatial relationship has received the most theoretical investigation^{5,7–12,19}. Here we focus on the temporal relationship in which the variance and mean abundance of a species through time are calculated for multiple populations, and these are graphed on a log–log scale with each datum representing a population. The null expectation for Taylor's power law for temporal variation is that the slope of the log variance versus log mean plot equals 2. This expectation derives from the well-known relationship that, when scaling any random variable X with finite mean μ and variance σ^2 by some constant k , the mean and variance of kX are $k\mu$ and $k^2\sigma^2$, respectively. On a log–log plot, the relationship between $k\mu$ and $k^2\sigma^2$ is a line with a slope of 2. A slope of less than 2 indicates that the per capita variability in population abundance decreases with increasing mean population abundance.

Taylor and others^{3,4} showed that the logarithm of the temporal variance of a population is tightly correlated to the logarithm of the mean abundance, with the correlation coefficient usually exceeding 0.9. For most species, the slope of the relationship is between 1 and 2. In an attempt to explain slopes of less than 2, Anderson and co-workers¹² showed that demographic stochasticity (variation caused by stochastic births and deaths of individuals in finite populations) in an exponentially growing population could lead to slopes for the temporal log variance versus log mean relationship of between 1 and 2, depending on the size of the population and the demographic parameters. Although demographic stochasticity certainly affects the dynamics of populations and can produce reduced slopes of Taylor's power law^{12,20,21}, these effects are most prevalent for population sizes on the order of hundreds or smaller. For many of the species considered by Taylor and others^{3,4}, the population size being sampled is likely to be in the tens of thousands or greater (for example, many insect species). For larger population sizes, environmental stochasticity in per capita population growth rates, which leads to a variance that scales with the square of the mean, overshadows the variance due to demographic stochasticity, which scales directly with the mean.

A second mechanism that could decrease the slope of Taylor's power law is sampling variance or error in the measurement of population abundance²². Sampling variance generally scales directly with the mean and thus by itself would produce a power-law slope of less than 2. Nevertheless, the effect of sampling variance on the slope declines with sampled abundance because at high abundance the variance–mean scaling will be dominated by environmental and/or demographic stochasticity. Although sampling variance certainly

contributes to the relationships seen in the empirical data, it is unlikely to be important in data sets where the sampled abundance is high.

Here we demonstrate that negative species interactions, including direct competition and 'apparent competition' through shared predators, lead to slopes of the log variance versus log mean relationship of between 1 and 2. Negative species interactions are common in natural communities²³, and thus this mechanism might affect the variance–mean scaling of many species. For example, power-law slopes of less than 2 have been demonstrated for numerous aphids in Europe^{3,4}, and many aphids are known to compete both directly²⁴ and indirectly²⁵. Certainly, for some species, a power-law slope of less than 2 is unlikely to be caused by negative species interactions, but many other species have direct competitors and shared predators that could affect their power-law slopes.

For the simplest case, we considered a competitive community of n species each growing according to a discrete logistic equation in which carrying capacities K_i differ between species. The density of species i at time $t + 1$, $x_i(t + 1)$, is given by

$$x_i(t + 1) = x_i(t) \exp \left[r_i \left(1 - \frac{x_i(t) + \sum_{j \neq i} \alpha_{ij} x_j(t)}{K_i} \right) + \varepsilon_i(t) \right] \quad (1)$$

where r_i is the intrinsic rate of increase of species i , α_{ij} is the

competition coefficient giving the per capita effect of species j on species i (which possibly differs between each pair of species) and $\varepsilon_i(t)$ is a normal random variable incorporating environmental stochasticity into the per capita population growth rate.

To simulate the data needed to calculate the temporal Taylor's power law, we embedded a focal species in different simulated communities. The focal species retained its intrinsic rate of increase, r_i , in each of these communities, but values of K_i and α_{ij} for the focal and all other species were selected from random distributions. One example of the resulting simulations (Fig. 1) with average $\alpha_{ij} = 0.25$ creates data corresponding to Taylor's power law with a slope of 1.58. Note that even though interspecific competition created a slope of less than 2, the strength of competition is weak (average $\alpha_{ij} = 0.25$). Furthermore, the average correlation coefficient between the focal species and all other species in the community (averaged over all five communities) is +0.04, giving no visual suggestion of a strong effect of competition on the temporal dynamics of the species.

Our simulations produce linear relationships between log variance and log mean comparable to those observed for a number of real species (Fig. 2a,b). When there are no species interactions (that is, $\alpha_{ij} = 0$), the slope of the temporal relationship between log variance and log mean averages 2 (Fig. 2c) but might be above or below 2 for a single set of simulations. However, when competitive interactions increase between species, the average slope decreases from 2 to 1 (Fig. 2c). Real single-species data sets show a range of slopes from 0.9 to 2.6, with most between 1.0 and 1.8 (ref. 3). Thus, the range of values described for real species could be the result of species existing in communities with strong to weak competition.

To explore the general relationships underlying these simulation

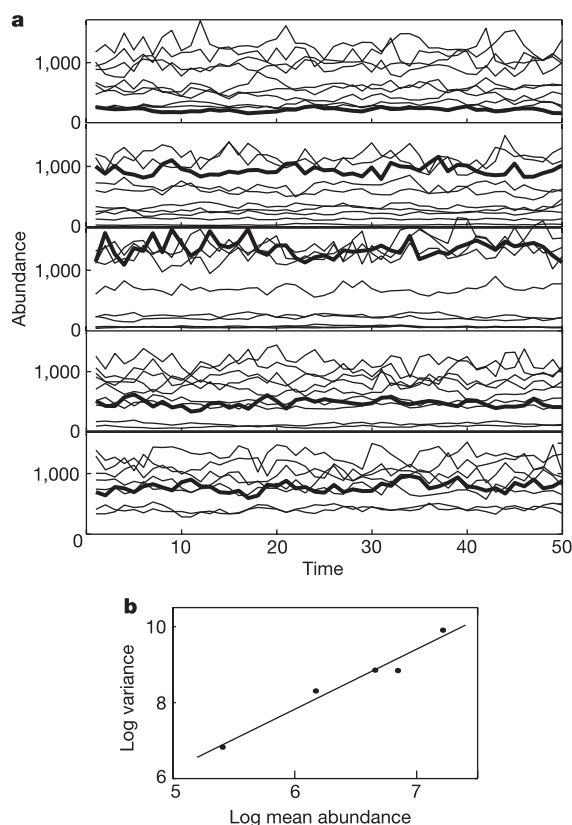


Figure 1 Five simulations of ten-species communities using the model from equation (1). **a**, Each species in the collection of simulations had a fixed r_i between 0.5 and 1.0 for all five simulations. In each simulation, corresponding to separate communities, the values of α_{ij} and K_j were selected from uniform distributions with ranges of 0–0.5 and 50–1,550, respectively. For the focal species, plotted with the thick solid line in each panel, $r = 0.57$ in all five communities. **b**, Plot of log variance against log mean for this species, with each point giving the log mean and log variance from one of the five other panels. The slope of the best-fit line is 1.58.

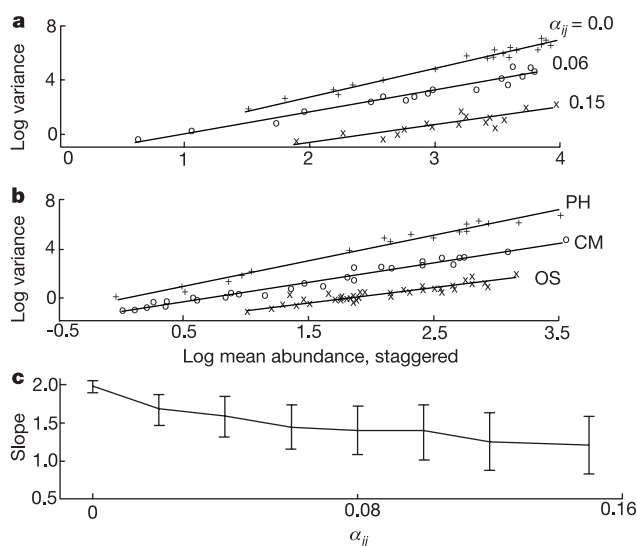


Figure 2 Taylor's power law for real and simulated data. **a**, **b**, Log temporal variance plotted against log mean abundance for simulations using equation (1) (**a**) and empirical data from three representative insect species from ref. 13 (**b**). PH, *Phorodon humuli*; CM, *Colostygia multistrigaria*; OS, *Ourapteryx sambucaria*. The mean abundances have been staggered horizontally for clarity. Parameter values from equation (1) were set to match the range in mean and variance for the data presented in **b**: plus signs, $\varepsilon \sim N(0, 0.006)$, $K \sim \text{uniform}(50-5,050)$; $\alpha_{ij} = 0$; circles, $\varepsilon \sim N(0, 0.006)$, $K \sim \text{uniform}(50-2,050)$, $\alpha_{ij} = 0.06$; crosses, $\varepsilon \sim N(0, 0.004)$, $K \sim \text{uniform}(50-1,050)$, $\alpha_{ij} = 0.15$. **c**, The slope (± 1 s.d.) of the plot of log temporal variance versus log mean abundance as a function of the average strength of competition between species, with $K \sim \text{uniform}(50-1,050)$. All simulations were done starting with 20 species, and temporal variance and mean were calculated from 50 time steps for each population after transients had dissipated.

results, we developed a log-linear approximation model that can be solved algebraically (see Methods). The solution incorporates the average strength of interspecific competition as the reduction in the average of the equilibrium densities of species in the community scaled by their carrying capacities. Thus, denoting $\bar{x}^* = \frac{1}{n} \sum_{i=1}^n \frac{x_i^*}{K_i}$, where x_i^* is the equilibrium abundance of species i (that is, the abundance that would occur in the absence of environmental variability), if there is no interspecific competition then $\bar{x}^* = 1$, and the greater the strength of interspecific competition averaged throughout the community, the lower is the value of $\bar{x}^*$. The analysis shows that the power-law slope decreases with increasing strength of interspecific competition, as measured by the reduction in $\bar{x}^*$, yet it is relatively insensitive to the number of species in a community once the community size, n , exceeds five (Fig. 3a). The extent to which competition decreases the slope also depends on how the stochastic variation affecting species densities is correlated: when the environmental stochasticity is perfectly correlated between species ($\rho = 1$), the slope equals 2, and decreasing environmental correlation decreases the slope (Fig. 3b). This shows, interestingly, that the slope of the relationship between log variance and log mean might be the product not only of species interactions but also of how species respond to environmental stochasticity relative to each other. Finally, higher intrinsic rates of increase lessen the effect of competition in decreasing the slope below 2 (Fig. 3b). These results hold for a broad suite of possible models of competitive interactions (see Methods). Furthermore, simulations of communities in which species do not interact directly but instead interact through shared predators show that indirect 'apparent competition' can produce power-law slopes of less than 2 in much the same way as direct competition does (results not shown).

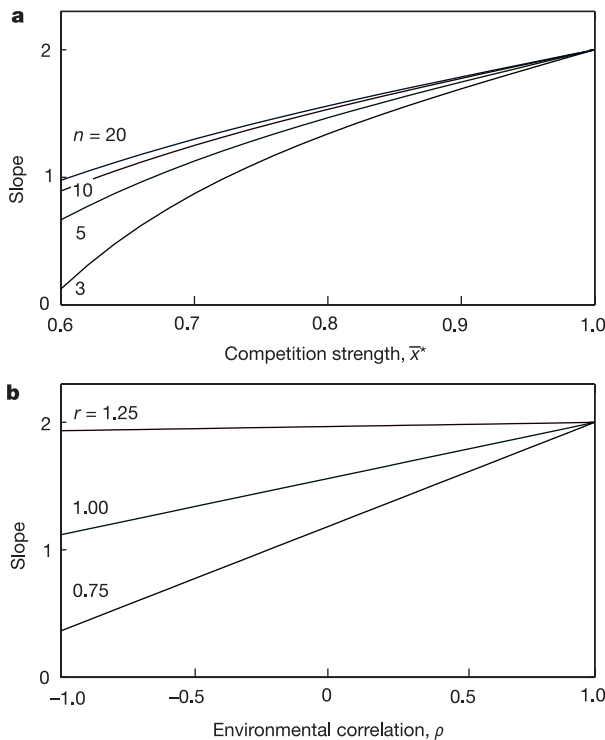


Figure 3 Analytical approximation of Taylor's power law (equations (5) and (6)). **a**, Slope of Taylor's power law plotted against the strength of competition measured by the decrease in mean species densities relative to their carrying capacities, $\bar{x}^*$, for communities with $n = 3, 5, 10$ and 20 species. Other parameters are $r = 1$ and $\rho = 0$. **b**, Slope of Taylor's power law plotted against the correlation between environmentally driven variation in species per capita population growth rates, ρ , for $r = 1.25, 1.00$ and 0.75 . Other parameters are $n = 20$ and $\bar{x}^* = 0.8$.

The effect of competitive interactions on the slope of Taylor's power law can be explained by considering the simple case of a community consisting of only two competitors. Suppose one competitor is less abundant, either because it has a smaller carrying capacity K_i or because it is an inferior competitor to the other species owing to asymmetric values of α_{ij} . The variability experienced by the rarer species is a combination of the environmental variability that affects it directly and the variability produced by competition with the more common species. Because the more common species has a higher carrying capacity and/or is a superior competitor, the variability in the rare species produced by interspecific competition is relatively large, and this substantially increases the variance in abundance of the rarer species. In contrast, the more common species experiences weaker interspecific competition and therefore has lower variability. Thus, the higher variability that the rarer species experiences through interspecific competition relative to the more common species produces a Taylor's power-law slope of less than 2. Although this example consists of two species in one community, the same reasoning applies to one species existing in two communities at different abundances.

In conclusion, our result contributes to the growing understanding that weak or diffuse interactions between species in a food web can contribute strongly to the dynamical properties of a community as a whole or its constituent species^{26–28}. If interspecific competition influences the slope of Taylor's power law for a given species, this species-specific slope is in fact governed by multi-species interactions. This raises caution about performing analyses involving the stochastic fluctuations in species population densities, such as population viability analyses¹³, because a species's stochastic dynamics might reflect not only the species's sensitivity to its environment but also its interactions with other species. Thus, to understand fluctuations in the abundance of individual species, it is necessary to understand how species interact with each other and how environmental variability is propagated through food webs. □

Methods

To analyse the competition model in equation (1), we first note that, without loss of generality, population densities, $x_i(t)$, can be rescaled relative to their carrying capacities, K_i , to give $\hat{x}_i(t) = x_i(t)/K_i$. Rescaling $\alpha_{ij} = \alpha_{ij}K_j/K_i$ gives an equation of the form of equation (1) but without parameters K_i . The rescaled coefficients $\hat{\alpha}_{ij}$ give the competitive effect of species j on species i weighted by the relative densities that each species would achieve in the absence of competition. Therefore, $\hat{\alpha}_{ij}$ measures the impact of interspecific competition relative to intraspecific competition.

Nearly all nonlinear models can be well approximated by Taylor expanding around a fixed point. For a model of n interacting species such as equation (1), the population dynamics can be approximated by the log-linear model

$$z_i(t+1) = \sum_{j=1}^n b_{ij}z_j(t) + \varepsilon_i(t) \quad (2)$$

where $z_i(t) = \log \hat{x}_i(t)$, $\varepsilon_i(t)$ is a random variable representing environmental stochasticity, and b_{ij} encapsulates the interactive effect of species j on the population growth rate of species i .

For the specific model in equation (1), we consider the special case in which $r_i = r$ and $\hat{\alpha}_{ij} = \hat{\alpha}_j$ for all species i . That is, all species have the same intrinsic rate of increase, and the relative competitive effect of species j on species i is the same for all species i . We consider this special case because it leads to an algebraic solution of equation (2). Numerical studies of the general form of equation (1) demonstrate the same qualitative patterns as derived from this special case. When $r_i = r$ and $\hat{\alpha}_{ij} = \hat{\alpha}_j$, coefficients in the log-linear approximation are

$$\left. \begin{aligned} b_{ij} &= 1 - r\hat{x}_i^* & (i=j) \\ b_{ij} &= b_j = -r \left(\frac{1 - r\bar{x}^*}{n-1} + \hat{x}_i^* \right) & (i \neq j) \end{aligned} \right\} \quad (3)$$

where $\hat{x}_i^*$ is the untransformed equilibrium abundance of species i (scaled by K_i) in the absence of environmental stochasticity, and $\bar{x}^*$ is the mean of the values of $\hat{x}_i^*$ across all species. This approximation has the properties that $b_{ij} = b_j$ and $b_{ii} + \sum_{j \neq i} b_{ij} = 1 - r$, which are properties shared by all competition models of the form

$$\hat{x}_i(t+1) = \hat{x}_i(t) f \left[\hat{x}_i(t) + \sum_{j \neq i} \hat{\alpha}_j \hat{x}_j(t) \right] \quad (4)$$

where $f[]$ is an unspecified linear or nonlinear function.

The change in the variance relative to the change in the mean can be computed from equations (2) and (3) as

$$\frac{\Delta V[z_i]}{\Delta \bar{z}_i} = \frac{(2n(n+1) - 6n + 2)(\rho - 1)(r(\bar{x}^*n - 1) - (n - 1))}{r(\bar{x}^*n - 1)((n - 2) + \bar{x}^*n - r(\bar{x}^*n - 1))(r(\bar{x}^*n - 1) - 2(n - 1))} \quad (5)$$

where $\Delta V[z_i]$ is the change in the variance of the log-transformed population density $z_i(t)$ with a change $\Delta \bar{z}_i$ in the mean, and ρ is the correlation of environmental stochasticity $\varepsilon_i(t)$ between species (assumed to be the same for all species in the community). This is an exact solution for the model given in equations (2) and (3) that can be derived algebraically for the cases of $n = 2$ and $n = 3$ (ref. 29), and then generalized to arbitrarily many species. The slope of the variance versus the mean population density on a logarithmic scale is

$$\frac{\Delta \log V[z_i]}{\Delta \log \bar{z}_i} \approx 2 + \frac{\Delta V[z_i]}{\Delta \bar{z}_i} \quad (6)$$

allowing the translation of equation (5) into Taylor's power law.

Received 20 September 2002; accepted 28 January 2003; doi:10.1038/nature01471.

1. Taylor, L. R. Aggregation, variance and the mean. *Nature* **189**, 732–735 (1961).
2. Taylor, L. R., Woivod, I. P. & Perry, J. N. The density-dependence of spatial behaviour and the rarity of randomness. *J. Anim. Ecol.* **47**, 383–406 (1978).
3. Taylor, L. R. & Woivod, I. P. Comparative synoptic dynamics: 1. Relationships between interspecific and intraspecific spatial and temporal variance-mean population parameters. *J. Anim. Ecol.* **51**, 879–906 (1982).
4. Taylor, L. R. & Woivod, I. P. Temporal stability as a density-dependent species characteristic. *J. Anim. Ecol.* **49**, 209–224 (1980).
5. Gillis, D. M., Kramer, D. L. & Bell, G. Taylor's Power Law as a consequence of Fretwell's ideal free distribution. *J. Theor. Biol.* **123**, 281–288 (1986).
6. Perry, J. N. Some models for spatial variability of animal species. *Oikos* **51**, 124–130 (1988).
7. Taylor, R. A. J. The behavioural basis of redistribution: 1. The DELTA-model concept. *J. Anim. Ecol.* **50**, 573–586 (1981).
8. Taylor, R. A. J. The behavioural basis of redistribution: 2. Simulations of the DELTA-model. *J. Anim. Ecol.* **50**, 587–604 (1981).
9. Perry, J. N. Chaotic dynamics can generate Taylor's power law. *Proc. R. Soc. Lond. B* **257**, 221–226 (1994).
10. Taylor, L. R. & Taylor, R. A. J. Aggregation, migration and population mechanics. *Nature* **265**, 415–421 (1977).
11. Hanski, I. Spatial patterns and movements in coprophagous beetles. *Oikos* **34**, 293–310 (1980).
12. Anderson, R. M., Gordon, D. M., Crawley, M. J. & Hassell, M. P. Variability and the abundance of animal and plant species. *Nature* **296**, 245–248 (1982).
13. Beissinger, S. R. & McCullough, D. R. (eds) *Population Viability Analysis* (Univ. Chicago Press, Chicago, 2002).
14. Ives, A. R. & Hughes, J. B. General relationships between species diversity and stability in competitive systems. *Am. Nat.* **159**, 388–395 (2002).
15. McCann, K. S. The diversity-stability debate. *Nature* **405**, 228–233 (2000).
16. Tilman, D. Biodiversity: Population versus ecosystem stability. *Ecology* **77**, 350–363 (1996).
17. Taylor, L. R. Synoptic dynamics, migration and the Rothamsted insect survey. *J. Anim. Ecol.* **55**, 1–38 (1986).
18. Green, R. H. Power analysis and practical strategies for environmental monitoring. *Environ. Res.* **50**, 195–206 (1989).
19. Thorarinnsson, K. Population density and movement: A critique of DELTA-models. *Oikos* **46**, 70–81 (1986).
20. Keeling, M. & Grenfell, B. Stochastic dynamics and a power law for measles variability. *Phil. Trans. R. Soc. Lond. B* **354**, 769–776 (1999).
21. Keeling, M. J. Simple stochastic models and their power-law type behaviour. *Theor. Popul. Biol.* **58**, 21–31 (2000).
22. Titmus, G. Are animal populations really aggregated? *Oikos* **40**, 64–68 (1983).
23. Schoener, T. W. Field experiments on interspecific competition. *Am. Nat.* **122**, 240–285 (1983).
24. Gianoli, E. Competition in cereal aphids (Homoptera: Aphididae) on wheat plants. *Environ. Entomol.* **29**, 213–219 (2000).
25. Muller, C. B. & Godfray, H. C. J. Apparent competition between two aphid species. *J. Anim. Ecol.* **66**, 57–64 (1997).
26. McCann, K., Hastings, A. & Huxel, G. R. Weak trophic interactions and the balance of nature. *Nature* **395**, 794–798 (1998).
27. Berlow, E. L. Strong effects of weak interactions in ecological communities. *Nature* **398**, 330–334 (1999).
28. Ives, A. R., Gross, K. & Klug, J. L. Stability and variability in competitive communities. *Science* **286**, 542–544 (1999).
29. Ives, A. R., Dennis, B., Cottingham, K. L. & Carpenter, S. R. Estimating community stability and ecological interactions from time-series data. *Ecol. Monogr.* (in the press).

Acknowledgements This manuscript was improved by the comments of B. Cardinale, K. Gross, L. Angeloni, A. Forbes and C. Williams. Funding was provided by the US National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.M.K. (e-mail: amkilpatrick@wisc.edu) or A.R.I. (e-mail: arives@wisc.edu).

Engineering evolution to study speciation in yeasts

Daniela Delneri^{*†}, Isabelle Colson^{*†‡}, Sofia Grammenoudi^{*†‡}, Ian N. Roberts[§], Edward J. Louis^{||} & Stephen G. Oliver^{*}

^{*} School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, UK

[§] National Collection of Yeast Cultures, Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, UK

^{||} Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

[†] These authors contributed equally to this work

The *Saccharomyces 'sensu stricto'* yeasts are a group of species that will mate with one another, but interspecific pairings produce sterile hybrids. A retrospective analysis of their genomes revealed that translocations between the chromosomes of these species do not correlate with the group's sequence-based phylogeny¹ (that is, translocations do not drive the process of speciation). However, that analysis was unable to infer what contribution such rearrangements make to reproductive isolation between these organisms. Here, we report experiments that take an interventionist, rather than a retrospective approach to studying speciation, by reconfiguring the *Saccharomyces cerevisiae* genome so that it is collinear with that of *Saccharomyces mikatae*. We demonstrate that this imposed genomic collinearity allows the generation of interspecific hybrids that produce a large proportion of spores that are viable, but extensively aneuploid. We obtained similar results in crosses between wild-type *S. cerevisiae* and the naturally collinear species *Saccharomyces paradoxus*, but not with non-collinear crosses. This controlled comparison of the effect of chromosomal translocation on species barriers suggests a mechanism for the generation of redundancy in the *S. cerevisiae* genome².

The *Saccharomyces sensu stricto* yeasts comprise six species: *S. cerevisiae*, *S. paradoxus*, *S. bayanus*, *S. cariocanus*, *S. mikatae* and *S. kudriavzevii*^{3,4}. Three of these species are characterized by the presence of specific reciprocal chromosomal translocations: four in *S. bayanus* and *S. cariocanus*, and one and two, respectively, for the two strains of *S. mikatae*—IFO1815 and IFO1816 (ref. 1). Reciprocal translocations in yeasts can potentially have a significant effect on reproductive isolation, and have been proposed as a cause of speciation, by inducing inviability of hybrid ascospores⁵. Yeasts from the *Saccharomyces sensu stricto* species readily hybridize in nature and in the laboratory, but hybrids are generally sterile, with less than 5% viable ascospores^{6–8}. We previously mapped the translocations, relative to the karyotype of *S. cerevisiae*, which are found in the *Saccharomyces sensu stricto* yeasts¹, and discovered that there was no correlation between the sequence-based phylogeny of these species and the occurrence of chromosomal translocations. We concluded that chromosomal translocations do not drive the process of speciation, but may contribute to the reproductive isolation between species once they had formed by some other means. To analyse that contribution, we required some way of dissecting the effects of reciprocal translocations away from the other genetic differences between these species. Fortunately, the availability of the complete genome sequence of *S. cerevisiae*⁹, and our own development of a method for generating precisely located chromosomal translocations in this organism⁹, offered the opportunity for effecting such a dissection. We could engineer the genome of *S. cerevisiae* to make it collinear with that of one of the other

[‡] Present addresses: School of Biological Sciences, University of Wales at Bangor, Brambell Building, Deiniol Road, Bangor LL57 2UW, UK (I.C.); Alexander Fleming Biomedical Science Research Centre, 14–16 Fleming Street, 16672-Vari, Greece (S.G.).

sensu stricto yeasts, and ask whether the two organisms were still reproductively isolated. In other words, were they still separate species?

We engineered chromosomes VI, VII and XVI of *S. cerevisiae* to produce the two reciprocal translocations found in the IFO1815 and IFO1816 strains of *S. mikatae*. These reconfigured *S. cerevisiae* strains were then used to assess the effect of reciprocal translocations in intraspecific and interspecific crosses. Fischer *et al.*¹ mapped the translocation breakpoint for *S. mikatae* IFO1816 between open reading frames (ORFs) YFR006w and YFR009w on chromosome VI, and between ORFs YGR021w and YGR026w on chromosome VII. We chose to engineer translocation breakpoints in the intergenic regions between these ORFs on each of the *S. cerevisiae* chromosomes. Diagnostic polymerase chain reaction (PCR) was used to confirm that we had obtained the intended two chromosomal products of reciprocal translocation in the engineered strain (see Fig. 1). The change in the karyotype was then verified by pulsed-field gel analysis, and a strain bearing the required translocation was retained and designated as Sct1. Recently, the genome sequence of *S. mikatae* 1816 has been released¹⁰, and we have confirmed that we had placed the VI/VII translocation breakpoints in *S. cerevisiae* within one and two ORFs, respectively, of their actual location in *S. mikatae* (see Methods and Supplementary Information for further details).

Sct1 was then manipulated to generate a second translocation between the right arm of chromosome VI (which comprises the genetic material of chromosome VII, because of the translocation introduced in Sct1) and the left arm of chromosome XVI. This would render it collinear with the second *S. mikatae* strain, IFO1815. For this construct, the translocation breakpoints were located between ORFs YGR188c and YGR189c, and YPL103c and YPL116w¹. In the case of this last breakpoint, the translocation falls in a rather large chromosomal interval (approximately 30 kilobases). No genomic sequence is available for 1815 and so, for technical reasons, we chose the largest intergenic region, between ORFs

YPL108w and YPL109c, to engineer the breakpoint in *S. cerevisiae*. This generated a new strain, Sct1/2, which is collinear to *S. mikatae* IFO1815. During the screen for Sct1/2, colonies carrying the second translocation (t2) but which had lost the original translocation (t1) from strain Sct1 (owing to a back-translocation between chromosomes VI and VII) were also isolated. Such a colony, carrying a translocation between the right arm of chromosome VII and the left arm of chromosome XVI, was retained and designated as Sct2. This strain differs from both *S. mikatae* IFO1815 and IFO1816 by one translocation and was then rendered kanamycin resistant by introducing the *KanMX* module into a phenotypically neutral telomeric locus, *AAD4* (ref. 11), in order to have a selectable marker to facilitate crosses with *S. mikatae* IFO1816 spores.

Hybrid zygotes carrying reciprocal translocations typically show meiotic anomalies and reduced fertility. The resulting levels of inviability depend on whether the translocated arms contain essential genes, and how the chromosomes segregate after recombination. To test the mating ability and fertility of the translocated strains Sct1, Sct2 and Sct1/2, we crossed them with wild-type *S. cerevisiae* and with the two *S. mikatae* strains (IFO1815 and IFO1816). To check the nature of the chromosomes present in the hybrid spores, we synthesized specific pairs of oligonucleotide primers for each of the 16 chromosomes of *S. cerevisiae* and *S. mikatae* (see Supplementary Table B). These primers were designed so that the PCR products deriving from the amplification of each chromosome of *S. mikatae* had a different length from those deriving from the homeologous chromosome of *S. cerevisiae*, thus enabling their separation by agarose gel electrophoresis. These analyses showed that, for every chromosome, there was at least one copy from each of the parental species, confirming the hybrid nature of the cells (Fig. 2). Although the quantitative validity of such PCR analyses is open to doubt, the data indicate that some of the hybrids may have been aneuploid at this early post-zygotic stage.

From the results of our isogenic crosses, it seems that the two engineered translocations affect hybrid sterility independently. The hybrids *S. cerevisiae* Sct1 × *S. cerevisiae* FY73, Sct2 × FY73 and *S. mikatae* IFO1816 × *S. mikatae* IFO1815 are heterozygous for one translocation, whereas the hybrid Sct1/2 × FY73 is heterozygous for two translocations. Hybrid Sct2 × FY73 showed a spore viability of 50.4%, whereas Sct1 × FY73 showed a spore viability of

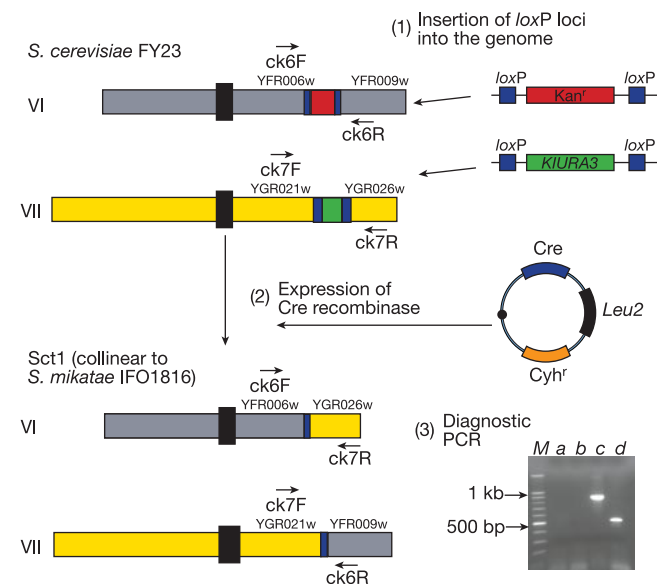


Figure 1 Strategy to create reciprocal translocations in yeast. The gene replacement technique was used to position the *loxP* loci on the desired chromosome (1). On induction of the *Cre* recombinase (2), the *loxP* sites recombine, provoking the pop-out of the selective markers and the creation of the reciprocal translocation. Diagnostic polymerase chain reaction (PCR) has been used to isolate the translocated colonies (3). In each PCR reaction, two pairs of primers were used: ck6F/ck6R (a), ck7F/ck7R (b), ck6F/ck7R (c) and ck7F/ck6R (d). M, 100-bp DNA ladder. Cyh^r, cycloheximide-resistant; Kan^r, kanamycin-resistant.

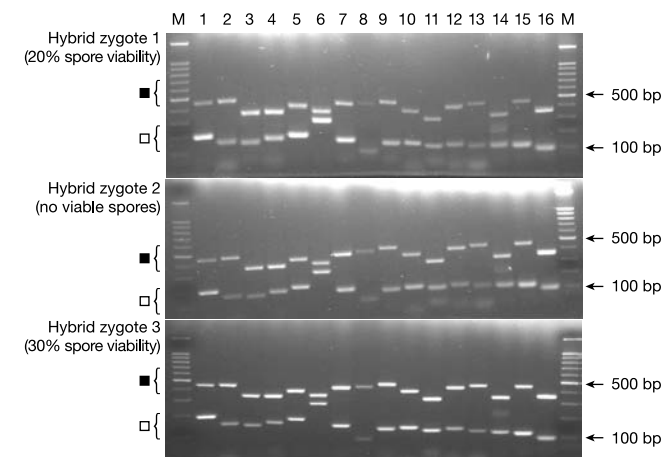


Figure 2 Discrimination between *S. cerevisiae* and *S. mikatae* chromosomes in hybrid zygotes. M, 100-bp DNA ladder; 1–16, PCR reactions using the DNA from hybrid zygotes (1, 2 and 3), as template, and four primers, two specific for each and every *S. cerevisiae* chromosome and two specific for *S. mikatae* chromosomes (chromosome I in lane 1, chromosome II in lane 2, and so on). *Saccharomyces cerevisiae* chromosomes are revealed by bands of sizes ranging between 350–450 bp (black boxes), whereas *S. mikatae* chromosomes are represented by bands of smaller sizes (100–250 bp; white boxes).

about 60%. Although an equal number of tetrads with four viable spores and with no viable spores was expected for the meiotic segregation of chromosomes differing by one translocation, the proportion of tetrads with four viable spores (33%) produced by the Sct1 × FY73 hybrid was significantly higher than the fraction of tetrads with no viable spores (24%). This segregation bias suggests that alternative segregation of translocation quadrivalents (leading to dyads with balanced chromosomes and four viable meiotic products) predominates over adjacent disjunction of the centromeres (leading to unbalanced dyads and hence four dead spores). Previous studies on eukaryotes had already shown that the 50% loss of fitness in translocants is somehow compensated, with the configuration of the quadrivalents at meiosis having a key role in the segregation of the four chromosomes^{12,13}. For the cross Sct1/2 × FY73, the expected spore viability of 25% was obtained. The cross between the two *S. mikatae* strains, which differ by one reciprocal translocation, led to a spore viability of about 40%. The results of all the crosses are summarized in Table 1.

Intraspecific crosses indicate that reciprocal translocations cause a marked reduction in hybrid fertility, and are therefore potential agents of reproductive isolation. But what is their contribution to the actual reproductive isolation between species? We used our translocated *S. cerevisiae* strains, together with wild-type strains of *S. cerevisiae*, *S. mikatae* and *S. paradoxus*, to assess the differential fitness of collinear or non-collinear interspecific crosses. Five to ten hybrids for each cross were tested for spore viability. In the crosses between *S. cerevisiae* and *S. mikatae*, the collinear crosses involved our engineered, translocated *S. cerevisiae* strains. However, in the crosses between *S. cerevisiae* and *S. paradoxus*, the engineered *S. cerevisiae* strain was used in the non-collinear cross. In both the *S. cerevisiae* × *S. mikatae* IFO1816 and the *S. cerevisiae* × *S. paradoxus* matings, none of the hybrids resulting from non-collinear crosses showed significant spore viability. However, four out of ten

hybrid zygotes analysed from the Sct1 × *S. mikatae* IFO1816 cross showed increased fertility, with spore viabilities ranging from 20% to 30%. Similarly, two out of five *S. cerevisiae* × *S. paradoxus* hybrids showed increased fertility (about 15%). The other cross, involving the doubly translocated Sct1/2 strain of *S. cerevisiae* and *S. mikatae*, failed to produce hybrids with increased fertility. This failure may have a variety of causes, including the more complex construction process involving two translocations, and differences in the selection of the hybrids. Another explanation could be the uncertainty about the breakpoints for the second translocation, which was not mapped as precisely as the first one. In this case, we could have incorporated genes into our engineered translocation that are not involved in the natural one in *S. mikatae*, therefore failing to restore total collinearity. In fact, small-scale chromosomal rearrangements, involving only a few genes, exist within the *Saccharomyces sensu stricto* species¹⁴, and these may cause a significant proportion of hybrid spores to be inviable¹⁵. Even when results from the Sct1/2 × *S. mikatae* hybrids are included in the comparison, collinear interspecific crosses have a significantly higher level of viability of their meiotic progeny than do hybrids from non-collinear interspecific crosses ($P = 0.0066$; Fisher exact test).

Fertile hybrids had not been seen in previous work with these species, where hybrids were generally selected by complementation of auxotrophic mutations or were isolated as single zygotes from mated pairs of cells^{3,6,7}. Although we cannot rule out the possibility that hybrid zygotes showing elevated fertility might arise from crosses between non-collinear strains, such events seem much less frequent. In cereals, crosses between divergent species have been made, but the success rate is very low. Fertile hybrid plants are usually aneuploid or polyploid, and translocations seem to increase their stability by reducing the frequency of recombination¹⁶. In yeast, it has been found that crosses between *S. cerevisiae* and *S. bayanus*, two species that differ by four translocations¹, could only very rarely produce tetrads with viable ascospores (at a frequency of about 10^{-4})¹⁷. Moreover, *S. pastorianus*, which is a hybrid of *S. cerevisiae* and *S. bayanus*¹⁸, hardly sporulates¹⁹. It seems

Table 1 Results of spore viability from collinear and non-collinear crosses

Collinear systems		Translocated systems	
Cross	Spore viability (%)	Cross	Spore viability (%)
Sc × Sc	90.3 (65/72)	Sct1 × Sc	60.4 (414/684)
Sct1 × Sct1	88.8 (71/80)	Sct2 × Sc	50.4 (121/240)
	96.1 (73/76)	Sct1/2 × Sc	24.6 (59/240)
Sm 1815 × Sm 1815	94.7 (72/76)	Sm 1815 × Sm 1816	41.2 (99/240)
Sm 1816 × Sm 1816	88.9 (64/72)		
Sct1 × Sm 1816	20 (138/692)	Sc × Sm 1816	0.6 (3/544)
	0.4 (1/272)		<1.7 (0/60)
	30.4 (95/312)		<1.7 (0/60)
	26.5 (53/200)		1.7 (1/60)
	1.8 (4/220)		1.7 (1/60)
	<0.4 (0/224)		3.3 (2/60)
	<1.2 (0/80)		<1.7 (0/60)
	<1.2 (0/80)		<1.7 (0/60)
	<1.2 (0/80)		<1.7 (0/60)
	16.6 (10/60)		
Sct1/2 × Sm 1815	<1.2 (0/80)	Sc × Sm 1815	0.8 (4/492)
	<1.4 (0/72)		1.7 (1/60)
	<1.2 (0/80)		<1.7 (0/60)
	1.2 (1/80)		1.7 (1/60)
	<1.2 (0/80)		<1.7 (0/60)
	<1.2 (0/80)		<1.7 (0/60)
		Sct × Sm 1815	1.1 (3/268)
			<1.3 (0/76)
			1.3 (2/160)
Sc × <i>S. paradoxus</i>	1.3 (2/160)	Sct1 × <i>S. paradoxus</i>	<1.2 (0/80)
	1.3 (2/160)		3.7 (3/80)
	15 (24/160)		<1.3 (0/80)
	20 (12/60)		2.6 (2/76)
	1.7 (1/60)		<1.3 (0/76)
			<1.3 (0/80)
<i>S. cariocanus</i> × <i>S. cariocanus</i>	92.1 (70/76)	Sc × <i>S. cariocanus</i>	0.6 (2/352)
			<1.2 (0/80)
			<1.2 (0/80)

Sc, *Saccharomyces cerevisiae*; Sm, *Saccharomyces mikatae*.

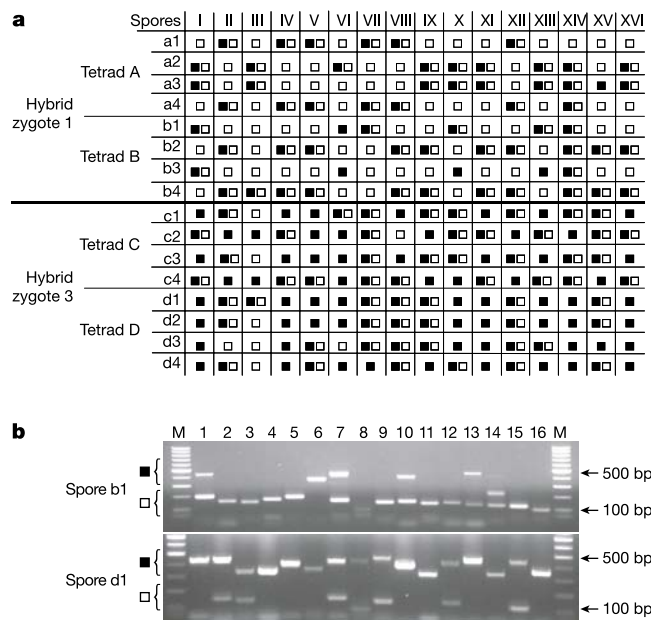


Figure 3 Diagnostic PCR for each of the 16 chromosomes of the two parent species reveals the extent of aneuploidy in the meiotic progeny of hybrid zygotes 1 (tetrads A and B) and 3 (tetrads C and D). Spores from hybrid zygote 1 have predominantly *S. mikatae* chromosomes (white boxes), whereas those from hybrid zygote 3 have mainly *S. cerevisiae* chromosomes (black boxes). **a**, Summary of karyotype data. **b**, Example PCR analyses for spores b1 and d1.

that the presence of translocations is preventing the generation of hybrid lines capable of producing a high proportion of viable spores.

The nature of the fertile interspecific hybrids produced by crossing collinear strains should be revealed by a study of their viable spore products. Viable spores resulting from the cross Sct1 × *S. mikatae* IFO1816 were characterized by a mis-segregation of visible markers (*kanMX*, *leu2*, *trp1*, *ura3*). PCR analyses were carried out, as before, to determine the parentage of the chromosomes. We analysed four sets of four viable spores deriving from the dissection of the hybrid zygotes 1 and 3 produced by the Sct1 × *S. mikatae* IFO1816 cross. The genomes of all spores were characterized by very widespread aneuploidy, involving nearly all chromosomes. Overrepresentation of *S. cerevisiae* chromosomes was observed in spores resulting from hybrid zygote 3; however, in spores resulting from the hybrid zygote 1, *S. mikatae* chromosomes predominated (Fig. 3). It is not known whether this aneuploidy occurred during meiosis or during the formation of the zygote itself; chromosome loss could have occurred during spore formation, or the zygotes themselves could carry unequal numbers of the homeologous chromosomes from each parent species. The observation that similar patterns of aneuploidy occur in different tetrads from the same hybrid clone (for example, tetrads A and B from hybrid zygote 1 in Fig. 3), as well as our PCR analyses, is most consistent with these hybrids being aneuploid at or soon after zygote formation, allowing for the production of viable spores. However, chromosome loss events could occur before, during, or after meiosis.

The pattern of spore karyotypes is compatible with extensive duplications of one of the parental species having occurred in the hybrid. This raises the possibility of genome duplication²⁰ by allopolyploidization or widespread aneuploidy in yeast. If the *Saccharomyces* genome was duplicated as a result of the hybridization of two species with collinear genomes, this could explain why no abrupt discontinuity in the sequence-based phylogeny of the yeasts has been discovered (data in ref. 21 and B. Dujon, personal communication). Recent studies have shown that the construction of allotetraploids by forced matings among the *S. cerevisiae sensu stricto* species results in hybrids with full spore viability²². Therefore the new hybrids we see here are not tetraploid. Nevertheless, the extensive aneuploidy that results from collinear interspecific crosses may be a major route to speciation, and it may explain the current 'patchwork' nature of the *S. cerevisiae* genome.

The principal problem in studying the molecular mechanisms of speciation is the difficulty in separating the effects of karyotypic rearrangements from genome-wide genetic incompatibilities, which undoubtedly have a major role in reproductive isolation^{5,23}. Uniquely, we have constructed strains differing by the presence of reciprocal translocations, but which were otherwise isogenic. Using these constructs, we have shown that chromosomal rearrangements do re-inforce reproductive isolation between *Saccharomyces* species. However, it is undeniable that genetic incompatibilities are a major reproductive isolation factor in present-day species. Indeed, many of the extant species have collinear genomes. Genetic divergence needs some sort of isolation (either geographical, ecological or behavioural) between incipient species to develop, and chromosomal rearrangements might have provided at least partial isolation. The availability of complete genome sequences for several members of the *Saccharomyces sensu stricto* group¹⁰ presents unparalleled opportunities for analysing the molecular mechanisms of reproductive isolation and species generation. Our study has provided both a conceptual and a methodological framework for such investigations, as it represents the first controlled comparison of the effect of chromosomal translocation on species barriers. □

Methods

Strains and medium

Saccharomyces paradoxus, *S. cariocanus*, *S. mikatae* IFO1816 and IFO1815 IC1 (*lys2*), *S.*

cerevisiae FY23 (*MATa*, *ura3-52*, *trp1Δ63*, *leu2Δ1*) and FY73 (*MATα*, *his3Δ200*, *ura3-52*) were grown on YPD medium (1% yeast extract, 2% peptone, 2% glucose), or minimal SD medium (0.67% yeast nitrogen base, 2% glucose) supplemented with any nutritional requirements, or YPGal medium (1% yeast extract, 2% peptone, 2% galactose).

For the selection of the kanamycin-resistant transformants, cells were grown on YPD plates containing 300 μg ml⁻¹ geneticin (G418; Gibco BRL). Hybrid zygotes (prototrophic and kanamycin-resistant) generated by crossing *S. cerevisiae* × *S. mikatae*, *S. cerevisiae* × *S. paradoxus* and *S. cerevisiae* × *S. cariocanus*, were selected on SD medium (0.67% yeast nitrogen base, 2% glucose) containing 0.24% urea as nitrogen source and 300 μg ml⁻¹ geneticin. For the selection of the cycloheximide-resistant transformants, the yeast cells were plated onto SD medium containing 2 μg ml⁻¹ cycloheximide (Boehringer Mannheim). Sporulation and presporulation medium, as well as all buffers and solutions, were made according to ref. 24.

Generation of chromosomal translocations

The translocations present in *S. mikatae* IFO1816 and IFO1815 were generated in *S. cerevisiae* by a Cre-mediated recombination event between two *loxP* loci inserted in the intergenic regions where the translocation points have been mapped^{1,11}. All the oligonucleotides used to create and verify the translocations between chromosomes VI and VII, and chromosomes VII and XVI are listed in Supplementary Table A, where a full description of the construction strategy and a discussion of the exact location of the translocation breakpoints used in relation to the genome sequence of *S. mikatae* 1816 (see ref. 10) may be found. Unless otherwise stated, all the PCR reactions for the amplification of the insertion cassettes (*loxP*-marker(s)-*loxP*) were performed at the annealing temperature of 55 °C and with an extension time of 2 min. Yeast transformation was performed using the lithium acetate method²⁵. The recombination between *loxP* sites located on different chromosomes was performed after transforming the yeast cells with a plasmid carrying a *Cre* recombinase gene under the control of the yeast *GAL1* promoter. *Saccharomyces cerevisiae* strains were screened for the presence of the correct translocations by colony PCR¹⁸ and by karyotypic analysis using pulsed-field gel electrophoresis²⁶.

S. cerevisiae × *S. mikatae* zygotes and tetrad analysis

Using a microneedle (Singer Instruments), single haploid cells of *S. cerevisiae* were put in physical contact with each of the spores coming from a single *S. mikatae* tetrad. In this way, two out of four spores were able to mate with *S. cerevisiae* cells of the opposite mating-type. Hybrids were selected on minimal medium containing geneticin (for hybrids *S. cerevisiae* kanamycin resistant × *S. mikatae* IFO1816), and on minimal medium only (for hybrids *S. cerevisiae* *MATa*, *ura3-52*, *trp1Δ63*, *leu2Δ1* × *S. mikatae* IFO1815 *IS1 lys2*). Tetrads were dissected with a Singer MSM micromanipulator.

Screening of the hybrid zygotes and spores

The presence of chromosomes belonging to *S. cerevisiae* and/or *S. mikatae* in the hybrid zygotes and spores was assessed by PCR using two pairs of diagnostic primers for each chromosome: one pair specific for the *S. cerevisiae* homeologue, and one pair specific for the *S. mikatae* homeologue (Supplementary Table B). The sequences of *S. mikatae* were retrieved from GenBank, and a BLASTn analysis versus the *S. cerevisiae* Genome Database (SGD) was performed to design species-specific primers (see Supplementary Table B). PCR reactions were performed at an annealing temperature of 57 °C, with an extension time of 30 s.

Received 21 November; accepted 31 December 2002; doi:10.1038/nature01418.

1. Fischer, G. *et al.* Chromosomal evolution in *Saccharomyces*. *Nature* **405**, 451–454 (2000).
2. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546–567 (1996).
3. Naumov, G. I. in *The Expanding Realm of Yeast-like Fungi* (eds de Hoog, G. S., Smith, M. T. & Weyman, A. C. M.) (Elsevier, Amsterdam, 1987).
4. Naumov, G. I., James, S. A., Naumova, E. S., Louis, E. J. & Roberts, I. N. Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *Int. J. Syst. Evol. Microbiol.* **50**, 1931–1942 (2000).
5. Ryu, S. L., Murooka, Y. & Kaneko, Y. Reciprocal translocation at duplicated *RPL2* loci might cause speciation of *Saccharomyces bayanus* and *Saccharomyces cerevisiae*. *Curr. Genet.* **33**, 345–351 (1998).
6. Naumov, G. I., Naumova, E. S. & Querol, A. Genetic study of natural introgression supports delimitation of biological species in the *Saccharomyces sensu stricto* complex. *System. Appl. Microbiol.* **20**, 595–601 (1997).
7. Marinoni, G. *et al.* Horizontal transfer of genetic material among *Saccharomyces* yeasts. *J. Bacteriol.* **181**, 6488–6496 (1999).
8. Seoighe, C. & Wolfe, K. H. Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol.* **2**, 548–554 (1999).
9. Delneri, D. *et al.* Exploring redundancy in the yeast genome: an improved strategy for the use of the *cre-loxP* system. *Gene* **252**, 127–135 (2000).
10. *Saccharomyces* Genome Database. (<http://genome-www4.stanford.edu/cgi-bin/FUNGI/FungiMap>).
11. Delneri, D., Gardner, D. C. J. & Oliver, S. G. Analysis of the seven-member AAD gene set demonstrates that genetic redundancy in yeast may be more apparent than real. *Genetics* **153**, 1591–1600 (1999).
12. Bengtsson, B. O. & Bodmer, W. F. The fitness of human translocation carriers. *Ann. Human Genet.* **40**, 253–257 (1976).
13. Loidl, J., Jin, Q.-W. & Jantsch, M. Meiotic pairing and segregation of translocation quadrivalents in yeast. *Chromosoma* **107**, 247–254 (1998).
14. Fischer, G., Neuvéglise, C., Durrens, P., Gaillardin, C. & Dujon, B. Evolution of gene order in the genomes of two related yeast species. *Genome Res.* **11**, 2009–2019 (2001).
15. Lynch, M., O'Hely, M., Walsh, B. & Force, A. The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**, 1789–1804 (2001).
16. Ramsey, J. & Schemske, D. W. Pathways, mechanisms, and rate of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* **29**, 467–501 (1998).

17. Sebastiani, F., Barberio, C., Casalone, E., Cavalieri, D. & Polsinelli, M. Crosses between *Saccharomyces cerevisiae* and *Saccharomyces bayanus* generate fertile hybrids. *Res. Microbiol.* **153**, 53–58 (2002).
18. Tamai, Y., Momma, T., Yoshimoto, H. & Kaneko, Y. Co-existence of two types of chromosome in the bottom-fermenting yeast, *Saccharomyces pastorianus*. *Yeast* **14**, 923–937 (1998).
19. Kiehlbrandt, M. C., Nilsson-Tillgren, T., Peterson, J. G. L., Hølemberg, S. & Gjermsen, C. in *Yeast Genetics* (eds Spencer, J. F. T., Spencer, D. M. & Smith, A. R. W.) 421–438 (Springer, New York, 1983).
20. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
21. Llorente, B. *et al.* Genomic exploration of the hemiascomycetous yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*. *FEBS Lett.* **487**, 122–133 (2000).
22. Greig, D., Borts, R. H., Louis, E. J. & Travisano, M. Epistasis and hybrid sterility in *Saccharomyces*. *Proc. R. Soc. Lond. B.* **269**, 1167–1171 (2002).
23. Hunter, N., Chambers, S. R., Louis, E. J. & Borts, R. H. The mismatch repair system contributes to meiotic sterility in an interspecific yeast. *EMBO J.* **15**, 1726–1733 (1996).
24. Adams, A., Gottschling, D. E., Kaiser, C. A. & Stearns, T. in *Methods in Yeast Genetics* (ed. Dickerson, M. M.) 145–175 (Cold Spring Harbor Laboratory Press, New York, 1997).
25. Gietz, R. D., Schiestl, R. H., Willems, A. R. & Woods, R. A. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* **11**, 355–360 (1995).
26. Gardner, D. C. J., Heale, S. M., Stateva, L. I. & Oliver, S. G. Treatment of yeast cells with wall lytic enzymes is not required to prepare chromosomes for pulsed-field gel analysis. *Yeast* **9**, 1053–1055 (1993).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements This work was supported by grants from the Biotechnology and Biological Sciences Research Council (to E.J.L., I.N.R. and S.G.O.) and the Wellcome Trust (to S.G.O.). We thank S. James and L. Lockhart for their help in some early analyses, and B. Dujon for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.G.O.
(e-mail: steve.oliver@man.ac.uk).

Direct measurement of the transfer rate of chloroplast DNA into the nucleus

Chun Y. Huang*, Michael A. Ayliffe† & Jeremy N. Timmis*

* Department of Molecular Biosciences, The University of Adelaide, South Australia, 5005, Australia

† CSIRO Plant Industry, GPO Box 1600, Australian Capital Territory 2601, Australia

Gene transfer from the chloroplast to the nucleus has occurred over evolutionary time¹. Functional gene establishment in the nucleus is rare, but DNA transfer without functionality is presumably more frequent. Here, we measured directly the transfer rate of chloroplast DNA (cpDNA) into the nucleus of tobacco plants (*Nicotiana tabacum*). To visualize this process, a nucleus-specific neomycin phosphotransferase gene (*neoSTLS2*) was integrated into the chloroplast genome, and the transfer of cpDNA to the nucleus was detected by screening for kanamycin-resistant seedlings in progeny. A screen for kanamycin-resistant seedlings was conducted with about 250,000 progeny produced by fertilization of wild-type females with pollen from plants containing cp-*neoSTLS2*. Sixteen plants of independent origin were identified and their progenies showed stable inheritance of *neoSTLS2*, characteristic of nuclear genes. Thus, we provide a quantitative estimate of one transposition event in about 16,000 pollen grains for the frequency of transfer of cpDNA to the nucleus. In addition to its evident role in organellar evolution, transposition of cpDNA to the nucleus in tobacco occurs at a rate that must have significant consequences for existing nuclear genes.

The genomes of cytoplasmic organelles have been depleted during endosymbiotic evolution, as many genes were either lost or transferred to the nucleus. The plastome of higher plants has been reduced to approximately 130 genes from an ancestral cyanobacterial genome that probably contained over 3,000 genes². It has been estimated that 1,700 protein-coding nuclear genes of *Arabidopsis thaliana* were acquired from cyanobacteria¹. Even greater genetic erosion has occurred during mitochondrial genome evolution³. The *rps10* gene, located in the mitochondrial genome in most angiosperms, has been transferred independently to the nucleus hundreds of times during endosymbiotic evolution⁴. Similar multiple independent gene transfers from the chloroplast to the nuclear genome have also been reported⁵.

Long tracts of extant organelle DNA are also found in the eukaryotic chromosomes of plants. For example, about 620 kilobases (kb) of mitochondrial DNA (mtDNA) are found on chromosome 2 of *Arabidopsis*⁶, and there are about 33 kb of cpDNA on chromosome 10L of rice⁷. Indeed, eukaryotes generally contain DNA sequences in their nuclear genome that show close similarity

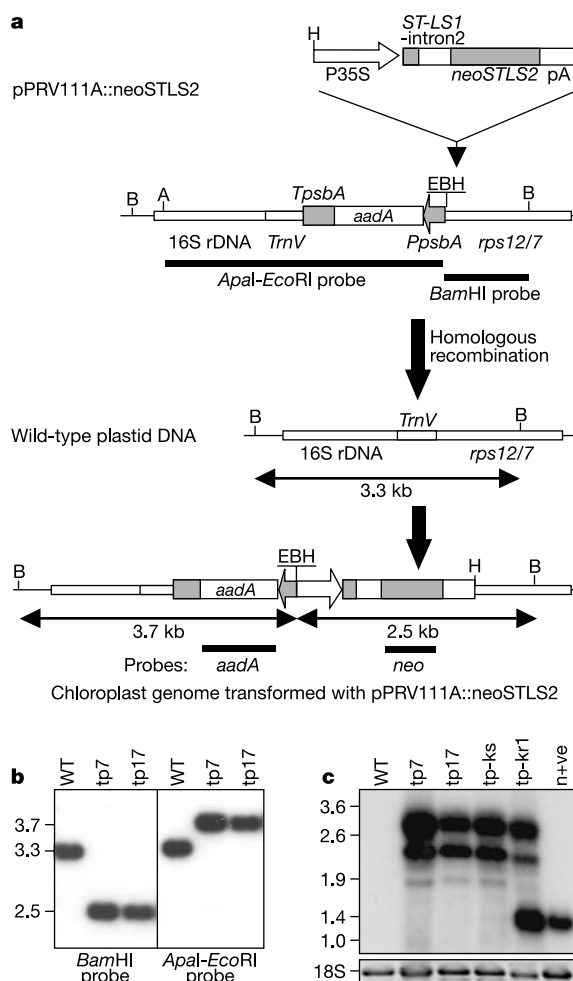


Figure 1 Introduction of *aadA* and *neoSTLS2* genes into the tobacco plastid genome by homologous recombination. **a**, Construction and integration of pPRV111A::neoSTLS2. Restriction sites are: A, *ApaI*; B, *BamHI*; E, *EcoRI*; H, *HindIII*. **b**, Southern analysis of two transplastomic lines. Total DNA (5 µg) was digested with *BamHI* and hybridized with the probes indicated (see **a**). **c**, Northern analysis of transplastomic lines and their progeny. Total leaf RNA (3 µg) was hybridized with the *neo* probe (see **a**). 18S RNA stained with ethidium bromide provides a loading control. Molecular sizes (in kilobases (kb)) are indicated on the left. n + ve, nuclear *neoSTLS2* control line.

to organellar DNA^{8–11}, suggesting that transfer of chloroplast and mtDNA are ongoing, possibly frequent processes.

Transfer of DNA from yeast mitochondria to the nucleus, in the form of a non-integrating recombinant plasmid, occurs at a rate of one in 2.5×10^4 cells per generation¹², but nothing is known about the frequency of transposition of cpDNA to the nucleus. To quantify this process, we inserted a nuclear-specific *neoSTLS2* gene (ref. 13) together with a plastid-selectable aminoglycoside 3'-adetyltransferase marker gene, (*aadA*)¹⁴, into the plastome of tobacco (Fig. 1a). The *aadA* gene confers spectinomycin resistance, enabling selection of transplastomic cells. The *neoSTLS2* gene features a constitutive plant viral promoter, CaMV-35S¹⁵, and a nuclear intron. It was designed to be functional only when transposed to the nucleus and to confer kanamycin resistance in young seedlings. An intron, from the potato nuclear *ST-LS1* gene¹⁶, was integrated within the reading frame to prevent synthesis of a functional protein in the chloroplast¹³.

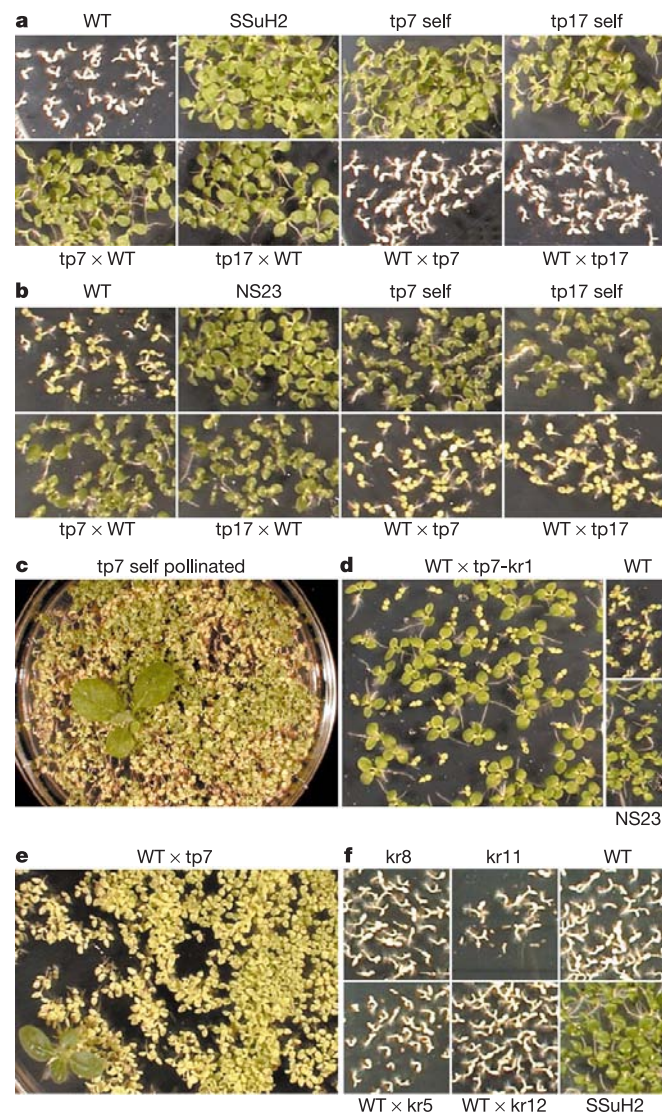


Figure 2 Seedling tests for resistance to spectinomycin and kanamycin. **a, b**, Seedlings from six progenies grown on medium containing $500 \mu\text{g ml}^{-1}$ spectinomycin (**a**) or $150 \mu\text{g ml}^{-1}$ kanamycin (**b**) with positive (SSuH2 or NS23) and wild-type controls. **c**, One kanamycin-resistant plant among the progeny of self-pollinated tp7 plants on kanamycin ($150 \mu\text{g ml}^{-1}$) and geneticin ($20 \mu\text{g ml}^{-1}$) medium. **d**, Segregation of kanamycin resistance in progeny of wild type $\times$ tp-kr1 plants on $150 \mu\text{g ml}^{-1}$ kanamycin medium. **e**, A kanamycin-resistant plant obtained in the progeny of wild type $\times$ tp7 cross. **f**, Progenies of four kanamycin-resistant plants with controls on spectinomycin medium.

Two independent transplastomic (tp) lines (tp7 and tp17), containing both *neoSTLS2* and *aadA*, were obtained by particle bombardment of tobacco leaves with pPRV111A::neoSTLS2 (Fig. 1a). DNA blot analysis confirmed the homoplasmic presence of the transgenes in tp7 and tp17 lines (Fig. 1b). Homoplasmy is not essential for our transposition experiments, but it may be critical for genetic stability of the transplastomic lines in the absence of spectinomycin selection.

Crosses were made to determine the inheritance and functionality of the selectable markers in the two transplastomic lines. Seeds set from tp7 or tp17 as the female parent were able to germinate on medium containing $500 \mu\text{g ml}^{-1}$ spectinomycin, and develop uniformly into green, healthy seedlings similar to a positive control, SSuH2 (ref. 17), containing a plastid *aadA* gene (Fig. 2a). In contrast, seedlings of wild-type tobacco and seedlings from wild-type females crossed with tp7 or tp17 as male parents, bleached rapidly on spectinomycin medium (Fig. 2a and Table 1). This reciprocal difference in the inheritance of *aadA* is characteristic of maternal inheritance of plastid genes in tobacco.

Notably, when the seeds set from tp7 or tp17 female parents were grown on medium containing $150 \mu\text{g ml}^{-1}$ kanamycin, all seedlings showed partial resistance to the antibiotic. They were uniformly paler green and smaller in size than positive control seedlings from a nuclear *neo* tobacco line, NS23 (ref. 18; Fig. 2b and Table 1). Wild-type seedlings and seedlings from wild-type females crossed with tp7 or tp17 male parents were all fully sensitive to kanamycin

Table 1 Progenies analysed for resistance to spectinomycin and kanamycin

Cross	Spectinomycin (500 µg ml ⁻¹)		Kanamycin (150 µg ml ⁻¹)		kr:ks ratio (P)
	r	s	r	s	
.....					
Section I					
tp7 self	168	0	0	827†	—
tp7 × WT	129	0	0	685†	—
WT × tp7	0	164	7	92,900	—
tp17 self	90	0	0	1,022†	—
tp17 × WT	127	0	0	532†	—
WT × tp17	0	90	11	165,000	—
Section II					
WT × tp-kr1	0	142	166	149	1:1 (NS)
kr1 self	ND	ND	140	34	3:1 (NS)
Section IIIA					
kr3 self	0	80	137	42	3:1 (NS)
kr6 self	0	38	107	35	3:1 (NS)
kr7 self	0	49	64	12	3:1 (NS)
kr8 self	0	82	165	49	3:1 (NS)
kr9 self	0	39	51	13	3:1 (NS)
kr10 self	0	35	62	17	3:1 (NS)
WT × kr12	0	79	129	120	1:1 (NS)
WT × kr13	0	100	133	133	1:1 (NS)
kr14 self	0	41	64	18	3:1 (NS)
kr16 self	0	48	126	60	3:1 (NS)
kr17 self	0	64	110	44	3:1 (NS)
kr18 self	0	40	52	15	3:1 (NS)
Section IIIB					
kr4 self	0	76	136	330	3:1 (*)
WT × kr4	0	84	80	72	1:1 (NS)
kr5 self	ND	ND	39	144	3:1 (*)
WT × kr5	0	73	72	76	1:1 (NS)
kr11 self	0	59	12	105	3:1 (*)
kr19 self	0	54	36	59	3:1 (*)
Section IIIC					
kr2 self	0	81	0	185	3:1 (*)
WT × kr2	ND	ND	0	224	1:1 (*)
kr15 self	0	39	0	162	3:1 (*)
WT × kr15	ND	ND	0	311	1:1 (*)

Progenies are grouped in sections according to the pattern of inheritance. Section III is subdivided to reflect genotypic and phenotypic differences. The parental origins of the 18 kanamycin-resistant plants in section III are not specified, as tp7 and tp17 were indistinguishable in every respect (Figs 1c, 2a, b and 3c, d). ND, not determined; NS, not significant; kr, kanamycin resistant; ks, kanamycin sensitive; r, resistant; s, sensitive; WT, wild type.

* $P > 0.01$ (χ^2).

†Partial resistance to $150 \mu\text{g ml}^{-1}$ kanamycin.

(Fig. 2b). These results indicate that the partial kanamycin-resistant phenotype is strictly maternally inherited, as expected for cp-*neoSTLS2* (refs 14, 19), and demonstrate that there is no functional *neoSTLS2* gene in the nucleus of tp7 or tp17 lines, which could have been introduced concomitantly during chloroplast transformation.

Two *neo*-hybridizing transcripts (3.3 and 2.2 kb) were identified in tp7 and tp17 RNA that were larger than correctly processed *neo* transcripts (1.2 kb) in RNA of a nuclear *neoSTLS2* control line, derived from pCMneoSTLS2 (ref. 13; Fig. 1c). However, a low level of the intron-spliced messenger RNA was detected in these two transplastomic lines by the more sensitive polymerase chain reaction with reverse transcription (RT-PCR) (data not shown). Therefore cp-*neoSTLS2* in the chloroplast gives rise to relatively abundant transcripts that, although inefficiently processed, probably account for the partial kanamycin resistance observed in the transplastomic lines and their progenies (Fig. 2b). To suppress this partial resistance, several analogues of kanamycin were tested. Supplementation of kanamycin medium with a low concentration of geneticin was found to be effective (Fig. 2c).

One kanamycin-resistant seedling was identified in an initial screen of 9,300 seeds from the progeny of self-pollinated tp7 on kanamycin and geneticin medium (Fig. 2c). This plant was designated tp-kr1 to reflect that it probably contained both chloroplast and nuclear copies of *neoSTLS2*. Abundant, correctly processed *neo* transcripts (1.2 kb) were identified in tp-kr1 RNA in addition to the 3.3- and 2.2-kb plastid transcripts (Fig. 1c). These 1.2-kb transcripts remained in kanamycin-resistant progeny after wild-type females

were crossed with tp-kr1, whereas the 3.3- and 2.2-kb chloroplast transcripts were absent owing to uniparental inheritance (data not shown).

To gain further evidence that the transferred cpDNA had integrated into the nucleus of tp-kr1, segregation analysis of kanamycin and spectinomycin resistance was applied to the progeny of wild type $\times$ tp-kr1 crosses. A 1:1 segregation ratio (kanamycin resistant/kanamycin sensitive) was observed (Fig. 2d and Table 1), demonstrating that at least one expressed copy of *neoSTLS2* had integrated into the nuclear genome of tp-kr1 at a single locus. Seedlings from this cross were uniformly sensitive to spectinomycin (Table 1), confirming non-transmission of the paternal transplastome. The kanamycin-resistant progeny of the wild type $\times$ tp-kr1 cross were designated kr1 to reflect the replacement of the transplastome with the wild type plastome. Self-fertilization of kr1 yielded a 3:1 segregation ratio (kanamycin resistant/kanamycin sensitive) in the progeny, demonstrating stable nuclear transmission of the transposed *neoSTLS2* gene (Table 1). A single fragment hybridized to *neo* and *aadA* probes in Southern hybridizations of *DraI*-digested kr1 DNA, consistent with a single cpDNA insert in nuclear DNA (data not shown). These data provide genetic and molecular evidence of a nuclear integrant containing *neoSTLS2* in tp-kr1. The unequivocal verification of nuclear integration provided by sequence analysis of a transplastomic/nuclear junction in kr1 is described later (see Fig. 3e).

More kanamycin-resistant individuals arising from independent transposition events were sought to quantify the frequency of cpDNA transfer by large-scale screening of seeds derived from

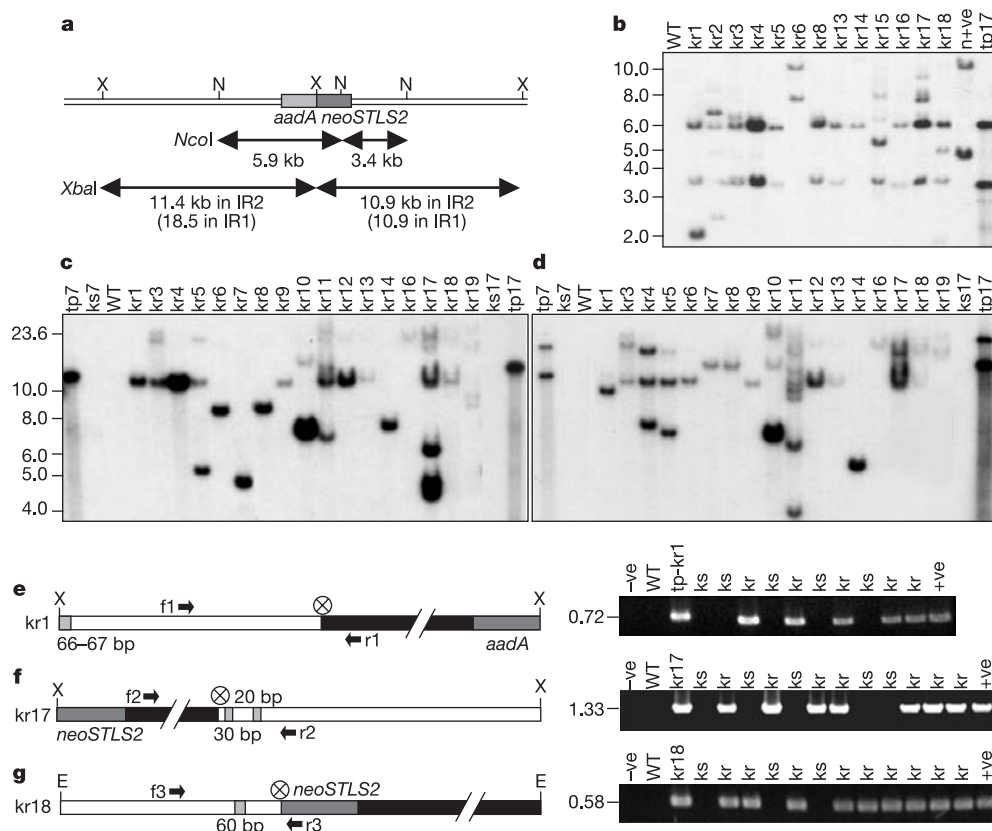


Figure 3 Analysis of kanamycin-resistant plants and three transplastomic/nuclear junctions. **a**, *NcoI* (N) and *XbaI* (X) transplastome fragments (double-headed arrows). **b**, *NcoI*-digested DNA from 13 kanamycin-resistant plants hybridized with *neo*. **c–d**, *XbaI*-digested DNA from 17 kanamycin-resistant plants hybridized with *neo* (**c**) or *aadA* (**d**). Lanes contained 5 μ g of DNA except for tp7 and tp17 (0.005 μ g). **e–g**, Unique transplastomic/nuclear junctions (circle enclosing a cross), PCR primers (arrows),

chloroplast sequences (filled boxes), unknown sequences (open boxes) and nuclear sequences with database homology (light grey boxes). Templates for PCR from the progeny of wild type $\times$ tp-kr1 (**e**), self-pollinated kr17 (**f**) or kr18 (**g**) are shown. –ve, no DNA; +ve, iPCR product. Gels were stained with ethidium bromide. Fragment sizes are indicated in kilobases. kr, kanamycin resistant; ks, kanamycin sensitive.

wild type $\times$ tp7 and wild type $\times$ tp17 backcrosses. Eighteen kanamycin-resistant plants (kr2–19) were identified among approximately 250,000 seedlings (Fig. 2e and Table 1). Backcross and self-pollinated progenies of all 18 kanamycin-resistant plants were uniformly sensitive to spectinomycin (Fig. 2f and Table 1), confirming the absence of pollen transmission of transplastomes from their paternal transplastomic lines. Kanamycin resistance segregated in the progenies of 16 kanamycin-resistant plants, 12 of which showed mendelian inheritance: either 3:1 (kanamycin resistant/kanamycin sensitive) after self-pollination or 1:1 after crossing to wild type (Table 1). Therefore at least one expressed copy of *neoSTLS2* is integrated into the nuclear genome at a single locus in these plants. The other four progenies of self-pollinated plants (kr4, kr5, kr11 and kr19) had segregation ratios that deviated significantly from 3:1. However, normal 1:1 ratios were restored in wild type $\times$ kr4 and wild type $\times$ kr5, the two backcross progenies tested (Table 1). Southern analysis identified multiple copies of *neoSTLS2* in these four plants, indicating that complex transgene interactions²⁰ may be responsible for the abnormal ratios observed in the progeny of self-pollinated plants (Fig. 3c).

Progenies of two kanamycin-resistant plants (kr2 and kr15) derived from self-pollination or backcrossing to wild-type females, showed uniform sensitivity to kanamycin (Table 1). Southern analysis confirmed that the *neoSTLS2* gene was present in both parent plants (Fig. 3b) but absent from any offspring (data not shown), indicating that the nuclear *neoSTLS2* gene failed to transmit. These data suggest that deletion may follow insertion at some chromosomal sites.

Total DNA samples from kr1 and 12 kanamycin-resistant plants from the large-scale screen were investigated to establish the independence of the transposition events. *NcoI* digestion and hybridization with the *neo* probe revealed two to four hybridizing fragments in each sample (Fig. 3a, b). All 13 kanamycin-resistant plants contained different sized fragments compared with the control nuclear plant. As *neoSTLS2* contains an *NcoI* site in its coding region, two *neo*-hybridizing fragments (3.4 and 5.9 kb) are expected in the transplastome (Fig. 3a). Ten out of the thirteen plant DNAs contained these two fragments, as did a low loading of DNA from tp17 (Fig. 3b). Therefore at least 9.3 kb of DNA, including *aadA*, *neoSTLS2* and adjacent cpDNA, was transferred from the transplastome and integrated into their nuclear genomes. These *aadA/neoSTLS2* integrants are flanked by at least 6.5 kb of cpDNA, which is larger than the plastid targeting sequence (3.0 kb) in the pPRV111A::neoSTLS2 transformation vector (Fig. 1a). This rules out the possibility that the *neo*-hybridizing fragments in most of the kanamycin-resistant plants could originate from any source other than the experimental transplastome.

To resolve further the independence of integration events, DNA samples from all kanamycin-resistant plants, except for kr2 and kr15, which did not transmit kanamycin resistance, were digested with *XbaI* for Southern analysis. The *neo* probe hybridized to one or more fragments, which differed in size and/or number, in each sample (Fig. 3c). Hybridization of the same membrane with the *aadA* probe similarly revealed fragments of different size and/or number in each sample (Fig. 3d). The patterns of hybridizing fragments for each DNA sample did not resemble that from any other kanamycin-resistant plant or of the parental transplastomic lines (Fig. 3c, d), indicating that each plant resulted from an independent integration event. No hybridization was observed in pooled DNA from 50 kanamycin-sensitive seedlings derived from wild type $\times$ tp7 or wild type $\times$ tp17 crosses (Fig. 3c, d), reconfirming that no *neo* or *aadA* sequences are present in the nuclear genome of the two transplastomic parental lines.

Junctions between three of the transposed chloroplast sequences and nuclear DNA were obtained (see Supplementary Information) by inverse polymerase chain reactions (iPCR). Analysis of a junction sequence from kr1 (1,315 bp) identified 1,128 bp of non-plastid

DNA adjoining the transposed cpDNA. A sequence located at the 5' end of this non-plastid DNA showed similarity to two nuclear encoded genes: 67 bp to a *Nicotiana plumbaginifolia* manganese superoxide dismutase gene (89%) and 66 bp to tobacco TA-29/TSJT1 (86%) (EMBL accession numbers Z67979 and X52283, respectively). No other significant homology was identified within this sequence, although a 300-bp open reading frame was present at the 3' end. To exclude the possibility that this iPCR product was artefactual, the junction sequence was amplified by PCR from genomic DNA. A 0.72-kb junction fragment was amplified from the parental tp7-kr1 DNA, DNAs from five kanamycin-resistant progeny of wild type $\times$ tp7-kr1 crosses and from the iPCR product (Fig. 3e). No PCR product was amplified from 5 kanamycin-sensitive progeny (Fig. 3e) or from the other 16 kanamycin-resistant plants derived from independent transpositions (data not shown). These results, together with the genetic and sequence data, unequivocally demonstrate that the transposed cpDNA in tp-kr1 is integrated into nuclear DNA.

Junction sequences from kr17 (1,451 bp) and kr18 (585 bp) DNA were shown to be unique to the nuclear genome of the plants from which they were derived and their kanamycin-resistant progeny (Fig. 3f, g). Sequence analysis of the non-plastid DNA adjacent to the kr17 and kr18 junctions showed no homology to sequences deposited in databases except for three short tracts (20–60 bp) that were similar to chloroplast DNA or mtDNA (Fig. 3f, g). As pre-existing cpDNA and mtDNA may be very common in the tobacco genome¹¹, it is not surprising to find other organellar DNA in the vicinity of an experimental integrant. The three junction sequences indicate that the transposed plastid fragments have integrated into different nuclear chromosomal locations.

These data demonstrate that DNA is transferred from the chloroplast and integrated into the nucleus at a frequency of one in approximately 16,000 tobacco pollen grains (16 independent, heritable nuclear insertions in 250,000 seedlings). This is a minimum estimate of the proportion of pollen grains that contain newly transposed cpDNA because our experimental approach would have identified only those integrants that were sufficiently large to contain an expressed *neoSTLS2* gene. In addition, as the integrants that we identified are smaller than the entire plastome, we expect similar transposition events involving other parts of the chloroplast genome that cannot be monitored by the *neoSTLS2* reporter gene.

Transfer of organellar DNA to the nucleus has been a major driving force in the evolution of eukaryotic cells. Transfer of cpDNA provides primary sequences for the evolution of new nuclear genes¹ and its frequency is such that integration could have a potentially wide impact on the function of native nuclear genes and genome organization. Despite this frequent transfer process the tobacco nuclear genome does not seem to be continuously expanding, so we assume that an equilibrium exists between integration and deletion. It is noteworthy that two kanamycin-resistant plants did not transmit the transposed plastid sequence to their progeny.

Transfer of cpDNA is the first evolutionary event necessary for functional establishment of a plastid gene in the nucleus. As the experimental cp-*neoSTLS2* gene contains a constitutive nuclear promoter, it allowed us to measure the primary rate of transposition to the nucleus, which does not reflect the problems faced by native chloroplast genes. The latter need to obtain appropriate nuclear sequences for transcription, translation and protein transport back into the chloroplast. A rate several orders of magnitude lower is expected for such complex DNA arrangements²¹.

The chloroplast is an increasingly attractive location for high level expression of prokaryote and eukaryote genes^{22,23}, and there is the advantage of natural biological containment afforded by maternal inheritance that prevents pollen-mediated transmission of the transplastome in many species^{19,24}. Our findings identify a potential pathway for transfer of plastid transgene DNA in tobacco. However, genes tailored for expression in the chloroplast are unlikely to

function in the nucleus as is demonstrated by the absence of spectinomycin resistance, despite the presence of a nuclear *aadA* gene in all of the kanamycin-resistant plants selected. □

Methods

pPRV111A::neoSTLS2 construction and transplastomic plants

pPRV111A::neoSTLS2 was produced by ligating the *Hind*III fragment containing neoSTLS2 from plasmid pCMneoSTLS2 (ref. 13) into pPRV111A¹⁴. This transformation vector was introduced into the plastids of *Nicotiana tabacum* L. Petit Havana (N,N) by biolistic bombardment, and homoplasmic plants were regenerated as described previously¹⁴. The homoplasmic plants were transferred to soil and grown in a controlled environment chamber with a 14 h light/10 h dark and 25 °C day/20 °C night growth regime. The photon flux density was approximately 300 $\mu\text{mol m}^{-2} \text{s}^{-1}$ at the plant surface. The nuclear neoSTLS2 control line was generated by biolistic bombardment with plasmid pCMneoSTLS2 (ref. 13).

Seedling tests for resistance to kanamycin and spectinomycin

Surface-sterilized seeds were plated on 150-mm plates containing 50 ml of 0.5 MS salt medium²⁵ and either spectinomycin dihydrochloride (500 $\mu\text{g ml}^{-1}$) or kanamycin sulphate (150 $\mu\text{g ml}^{-1}$). For screening seeds of self-pollinated tp7, the latter medium was supplemented with geneticin disulphate (20 $\mu\text{g ml}^{-1}$) when seedlings were two weeks old. The plates were placed at 25 °C with continuous fluorescent light. To determine nondestructively the kanamycin-resistance phenotype for verification of transplastomic/nuclear junctions, leaf pieces were taken from individual seedlings grown in 0.5 MS medium without kanamycin, and then cultured in MS medium containing 150 $\mu\text{g ml}^{-1}$ kanamycin and the required plant hormones²⁵. Kanamycin resistance was judged by callus growth.

Molecular analysis

DNA and RNA blot analyses were carried out as described^{11,26}. Junction sequences were obtained using iPCR as described (see <http://arabi4.agr.hokudai.ac.jp/ArabiE/protocols/general/general.html>). Key differences in the restriction fragment sizes between nuclear and transplastomic DNA were used to design cpDNA primers adjacent to the site of integration. These primers were used in conjunction with a second primer specific for either *aadA* or *neoSTLS2*. Primers used in iPCR were 5'-GAAGTTTCCAAAAGTCTGTT-3' with 5'-CTCGCATCTATTTTCATTG-3' for kr1, 5'-CCAGATTCCAAATGAACAAA-3' (f2) with 5'-CAATAGCCCTCTGGTCTCT-3' (r3) for kr17, and f2 with r3 for kr18. PCR amplification¹¹ of the junction sequences was undertaken with primers 5'-GCACGTGTGTCATTCAATACT-3' (f1) and 5'-CCAATTGTGACATCCCTCT-3' (r1) for kr1, f2 and 5'-GGTTTCCAAAGGGGTTT-3' (r2) for kr17, and 5'-GCCGTCATCAC TAACCATT-3' (f3) and r3 for kr18.

Received 26 November 2002; accepted 13 January 2003; doi:10.1038/nature01435.

Published online 5 February 2003.

1. Martin, W. *et al.* Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl Acad. Sci. USA* **99**, 12246–12251 (2002).
2. Kaneko, T. *et al.* Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109–136 (1996).
3. Gray, M. W. The endosymbiont hypothesis revisited. *Int. Rev. Cytol.* **141**, 233–357 (1992).
4. Adams, K. L., Daley, D. O., Qiu, Y. L., Whelan, J. & Palmer, J. D. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* **408**, 354–357 (2000).
5. Millen, R. S. *et al.* Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* **13**, 645–658 (2001).
6. Stupar, R. M. *et al.* Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats. *Proc. Natl Acad. Sci. USA* **98**, 5099–5103 (2001).
7. Yuan, Q. *et al.* Genome sequencing of 239-kb region of rice chromosome 10L reveals a high frequency of gene duplication and a large chloroplast DNA insertion. *Mol. Genet. Gen.* **267**, 713–720 (2002).
8. Timmis, J. N. & Scott, N. S. Spinach nuclear and chloroplast DNAs have homologous sequences. *Nature* **205**, 65–67 (1983).
9. Mourier, T., Hansen, A. J., Willerslev, E. & Arctander, P. The human genome project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol. Biol. Evol.* **18**, 1833–1837 (2001).
10. Ricchetti, M., Fairhead, C. & Dujon, B. Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* **402**, 96–100 (1998).
11. Ayliffe, M. A. & Timmis, J. N. Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA. *Theor. Appl. Genet.* **85**, 229–238 (1992).
12. Thorsness, P. E. & Fox, T. D. Escape of DNA from the mitochondria to the nucleus in the yeast, *Saccharomyces cerevisiae*. *Nature* **346**, 376–379 (1990).
13. Maas, C. *et al.* Expression of intron modified *NPTII* genes in monocotyledonous and dicotyledonous plant cells. *Mol. Breed.* **3**, 15–28 (1997).
14. Zoubenko, O. V., Allison, L. A., Svab, Z. & Maliga, P. Efficient targeting of foreign genes into the tobacco plastid genome. *Nucleic Acids Res.* **22**, 3819–3824 (1994).
15. Benfey, P. N. & Chua, N. H. The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. *Science* **250**, 959–966 (1990).
16. Eckes, P., Rosahl, S., Schell, J. & Willmitzer, L. Isolation and characterization of a light-inducible, organ-specific gene from potato and analysis of its expression after tagging and transfer into tobacco and potato shoots. *Mol. Gen. Genet.* **205**, 14–22 (1986).
17. Whitney, S. M. & Andrews, T. J. The gene for the ribulose-1, 5-bisphosphate carboxylase/oxygenase

- (Rubisco) small subunit relocated to the plastid genome of tobacco directs the synthesis of small subunits that assemble into Rubisco. *Plant Cell* **13**, 193–205 (2001).
18. Horser, C., Abbott, D., Wesley, V., Smith, N. & Waterhouse, P. Gene silencing—principles and application. *Genet. Eng. Principles Meth.* **24**, 239–256 (2002).
19. Daniell, H., Datta, R., Varma, S., Gray, S. & Lee, S. B. Containment of herbicide resistance through genetic engineering of the chloroplast genome. *Nature Biotechnol.* **16**, 345–348 (1998).
20. Fagard, M. & Vaucheret, H. (Trans)gene silencing in plants: how many mechanisms? *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 167–194 (2000).
21. Babiychuk, E., Fuangthong, M., van Montagu, M., Inzé, D. & Kushnir, S. Efficient gene tagging in *Arabidopsis thaliana* using a gene trap approach. *Proc. Natl Acad. Sci. USA* **94**, 12722–12727 (1997).
22. Maliga, P. Engineering the plastid genome of higher plants. *Curr. Opin. Plant Biol.* **5**, 164–172 (2002).
23. Daniell, H., Khan, M. S. & Allison, L. Milestones in chloroplast genetic engineering: an environmentally friendly era in biotechnology. *Trends Plant Sci.* **7**, 84–91 (2002).
24. Scott, S. E. & Wilkinson, M. J. Low probability of chloroplast movement from oilseed rape (*Brassica napus*) into wild *Brassica rapa*. *Nature Biotechnol.* **17**, 390–392 (1999).
25. Murashige, T. & Skoog, F. A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant* **15**, 473–497 (1962).
26. Huang, C. Y., Barker, S. J., Langridge, P., Smith, F. W. & Graham, R. D. Zinc deficiency up-regulates expression of high-affinity phosphate transporter genes in both phosphate-sufficient and -deficient barley roots. *Plant Physiol.* **124**, 415–422 (2000).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank the Australian Research Council for financial support; P. Maliga for pPRV111A; S. Whitney and J. Andrews for facilities and advice on chloroplast transformation, for SSuH2 seeds and for comments on the manuscript; T. J. Higgins, P. Whitfield, W. Martin and R. Lockington for discussion; C. Maas for pCMneoSTLS2; N. Smith for the NS23 seeds; and S. Elts for technical assistance.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.N.T. (e-mail: jeremy.timmis@adelaide.edu.au). Nucleotide sequences of three transplastomic/nuclear junctions are deposited in EMBL under accession numbers AJ495859 for kr1, AJ517467 for kr17 and AJ517468 for kr18.

Optimal transsaccadic integration explains distorted spatial perception

Matthias Niemeier^{*†‡}, J. Douglas Crawford^{†‡} & Douglas B. Tweed^{*†‡}

^{*} Departments of Physiology and Medicine, University of Toronto, 1 King's College Circle, Toronto M5S 1A8, Canada

[†] Canadian Institutes of Health Research, Group for Action and Perception

[‡] Centre for Vision Research, York University, 4700 Keele Street, Toronto M3J 1P3, Canada

We scan our surroundings with quick eye movements called saccades, and from the resulting sequence of images we build a unified percept by a process known as transsaccadic integration. This integration is often said to be flawed, because around the time of saccades, our perception is distorted^{1–6} and we show saccadic suppression of displacement (SSD): we fail to notice if objects change location during the eye movement^{7,8}. Here we show that transsaccadic integration works by optimal inference. We simulated a visuomotor system with realistic saccades, retinal acuity, motion detectors and eye-position sense, and programmed it to make optimal use of these imperfect data when interpreting scenes. This optimized model showed human-like SSD and distortions of spatial perception. It made new predictions, including tight correlations between perception and motor action (for example, more SSD in people with less-precise eye control) and a graded contraction of perceived jumps; we verified these predictions experimentally. Our results suggest that the brain constructs its evolving picture of the world by optimally integrating each new piece of sensory or motor information.

We make two or three saccades per second, collecting a rapid series of visual snapshots of the world. To perceive a large, changing scene—for example, when navigating through traffic—we mentally join up the images like pieces of a spatiotemporal jigsaw puzzle. To work out which piece of the puzzle goes where, the brain uses information from three main sources, all of them imperfect. One source is the retinal locations of images, but retinal acuity declines with eccentricity^{9,10}. The second source is the velocity signals from visual motion detectors; these are accurate when the eyes are stationary^{11,12}, but ineffective during saccades^{13–16}. The third is our sense of eye position, derived from muscle spindles¹⁷ or motor commands^{18–20}, but this information is also imperfectly reliable, especially during saccades^{21,22}.

Is it possible that transsaccadic integration uses its limited inputs in an optimal way, despite apparent flaws such as SSD? To see what optimal integration would look like, we simulated a visuomotor system that receives these three signals. It makes saccades, and

sometimes during or between those saccades, an object in its visual field jumps to a new location. The system judges whether a jump has occurred, and estimates its size and direction. For its deductions, the system knows the probability distributions of its unreliable signals, and of its own saccades and of jumps in the world. It makes optimal use of all its information to deduce the events in its surroundings (see Methods).

We simulated the model's performance in a typical psychophysical experiment: guessing in which of two possible directions a spot has jumped. Figure 1a plots the model's probability of choosing correctly as a function of jump size. If the spot jumps when the eye is stationary, the model (dotted line) notices: its perceptual threshold, defined as the smallest jump whose direction it can guess correctly 75% of the time, is very low. If the spot jumps during a saccade, the threshold is higher⁷ and it increases with saccade amplitude⁸. As shown in the inset, this increase is approximately linear, also in agreement with published data^{23,24}. Our own subject in Fig. 1b, given the same tasks, shows the same behaviours. Thus, optimal transsaccadic integration results in SSD, and in a human-like way. So, SSD need not reflect flawed transsaccadic integration, but is an unavoidable consequence of optimal inference from imperfect signals.

To test this model further, we used it to generate new predictions about SSD. The model shows SSD for two reasons, the first being that it attributes apparent jumps to errors in its visual and motor signals; therefore, it will show more SSD the more uncertain it is about its sensorimotor data. If its eye-position sense, like that of humans²¹, is less reliable parallel to a saccade than orthogonal to it, then it will be more blind to target jumps in the parallel dimension. This behaviour is shown in Fig. 1c, in which the model senses smaller jumps orthogonal to its saccades (dashed line) than parallel to them (solid line). To test this prediction experimentally, we had seven people perform the same tasks (see Methods). Like the typical subject in Fig. 1d, all subjects matched the model's predictions: they showed much less SSD for orthogonal than for parallel jumps ($P = 0.004$).

Also according to the model, SSD in any direction should depend on the combined uncertainty of retinal localization and eye-position sense in that direction. We can estimate this uncertainty by measuring the scatter (the standard deviation, s.d.) of eye positions right after the saccade (see Methods); we would expect a person's SSD to correlate with the postsaccadic scatter of their eye positions. The model predicts tighter correlations if we 'cancel out' other intersubject differences by taking ratios, as shown in Fig. 2a. Here, the variable on the abscissa is the 'eye-position scatter ratio'; for example, if after saccades your eye positions are spread out three times as much in the direction of the saccade than in the orthogonal

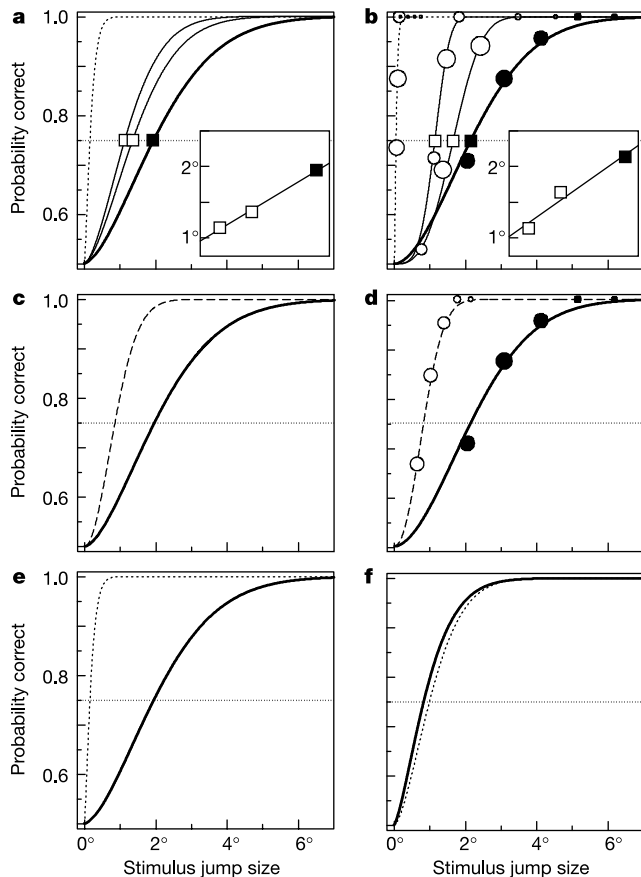


Figure 1 Predicted and actual saccadic suppression of displacement (SSD). **a**, The model judges in which of two directions a spot has jumped. If the spot jumps when the eye is fixating (dotted line), the threshold of perception (the jump size at which the model guesses the direction correctly 75% of the time) is just 0.2°. During 7.5° saccades (thin solid line), the threshold increases to 1.1°; during 10° saccades (medium solid line), it rises to 1.4°; and with 15° saccades (thick solid line), to 2°. The inset plots thresholds as a function of saccade size. **b**, A human subject's judgements on the same tasks. **c**, In the model, thresholds are lower for jumps orthogonal to the saccade (dashed line) than for parallel jumps (solid line). **d**, On the same tasks, a typical subject shows the same behaviour. Areas of circles represent the numbers of trials for each jump size. **e**, Simulating the blanking effect. When the target is continuously presented, the model shows low thresholds during fixation (dotted line) and high thresholds during saccades (solid line). **f**, When the target is blanked, the thresholds shift in opposite directions.

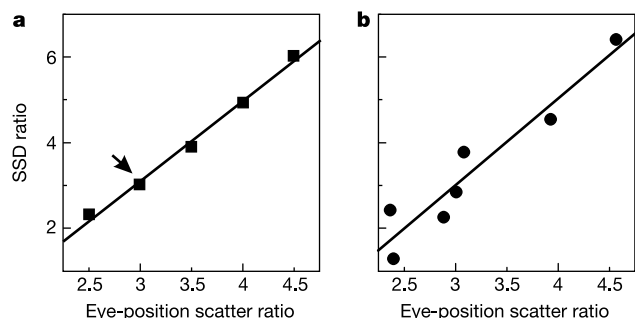


Figure 2 Perception and motor control. **a**, The optimal-integration model predicts a monotonic, roughly linear relationship between sensorimotor uncertainty, as reflected in eye-position scatter ratios, and SSD ratios (see text). **b**, Data from seven subjects confirm the prediction. Lines in both plots are best-fit linear regressions.

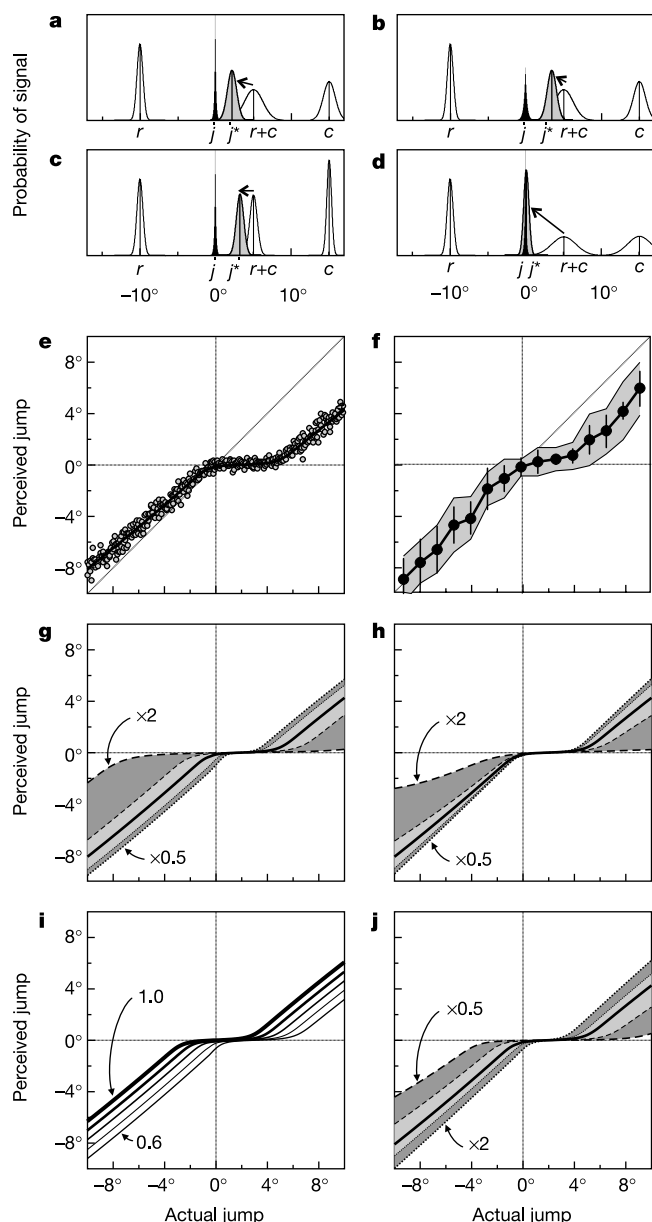


Figure 3 Contraction of perceived jumps. **a**, The percept j^* reflects a tug of war between the prior probability of a jump, j , and the sensorimotor estimate $r + c$, where r is the retinal signal and c is the efference copy signal; during saccades, v (an independent estimate derived from visual velocity detectors of the change in retinal location) has no role. **b, c**, j^* stays closer to $r + c$ when the prior probability of non-zero jumps is higher (that is, when $p(j)$ is wider; **b**), or when r and c are less variable (and therefore more reliable; **c**). **d**, But when r and c are highly variable, j^* gravitates strongly towards zero. **e**, The model predicts contraction: the slope of the curve is shallow near the middle. Grey circles are 1,000 responses, after 15° saccades. The solid line indicates average model performance; the straight diagonal line, ideal perception. **f**, All eight subjects showed similar contraction; this graph shows their bin-wise averaged responses, with standard deviation (s.d.) bars and averages of the individual s.d.s (grey envelope). Remaining panels illustrate the sensitivity of j^* to the model's main parameters. **g**, Influence of eye-position uncertainty, σ_c . The thick solid curve is the model's performance when $\sigma_c = 2.5(0.01d + 0.05^\circ)$, where d is the target distance (see Methods). Other curves show the results when this σ_c is multiplied by 2 and 1.3 (dashed lines), and by 0.7 and 0.5 (dotted line). **h**, Retinal-shift uncertainty, σ_r . Thick curve: $\sigma_r = 0.02e + 0.15^\circ$, where e is the postsaccadic, or postjump, eccentricity of the target. Other curves: σ_r multiplied by 2, 1.3, 0.7 and 0.5. **i**, Underestimation of saccade size, D . With more pronounced underestimation, the j^* curve shifts in the direction of the saccade. From left to right, $D = 1, 0.9, 0.8, 0.7$ and 0.6 . **j**, Width of the probability density function $p(j)$. The thick curve is the model's performance when the width parameter, w , is 0.065° . For the other curves, w was multiplied by 0.5, 0.7, 1.3 and 2.

direction, then your scatter ratio is 3, as it is for the simulated person marked with an arrow in the figure. On the ordinate is the 'SSD ratio': if your threshold for sensing a target jump is three times as large when the jump is parallel to your saccade than when it is orthogonal, then your SSD ratio is 3, as it is for the same data point in Fig. 2a. The key prediction, shown in the graph, is that people with higher scatter ratios should have higher SSD ratios: the latter ratio should rise as a function of the former, monotonically and roughly linearly.

We tested this prediction on seven people. The results, given in Fig. 2b, confirm the model's predictions: subjects with larger scatter ratios had larger SSD ratios, with a correlation of $r = 0.95$ ($P < 0.01$). The human data in Fig. 2b followed a rising, roughly linear curve like those of the optimal integrators in Fig. 2a.

To generate further tests, we turned to the second reason the model shows SSD: namely, the sheer implausibility, outside a vision laboratory, of an otherwise stationary object jumping in perfect synch with a saccade. Because of this prior improbability, the optimal integrator distrusts its inputs when they signal a jump. So, the model implies that SSD should diminish when the context makes jumps more plausible. For example, it concurs with the finding of Gysen *et al.*²⁵ that SSD is reduced if the object is already moving before it jumps. And it fits Deubel and colleagues' observation²⁶ that SSD is reduced when the jumping object is blanked for a time before it reappears in its new location (blinking makes it plausible that the postsaccadic spot is a new object in a new location, or that the object passed behind a barrier and re-emerged later, both of which scenarios are more likely than an abrupt translocation). To simulate this, we programmed our model to regard apparent jumps as more plausible when they coincided with brief disappearances. The simulations (Fig. 1e–f) mimic the reduced SSD reported by Deubel *et al.*²⁶, and their observation that blanking makes jumps less perceptible when the eye is stationary (because motion cues are removed).

Previous accounts of SSD have considered only two-alternative tasks in which subjects detect jumps or judge their directions, but the optimal-integration model also predicts a graded contraction of perceived jumps. The mechanism is illustrated in Fig. 3a–d. In Fig. 3a, the efference copy signal, c , reports a saccade of 15° while the retina, r , reports a stimulus shift of -10° , although both reports carry some uncertainty, as represented by the gaussian distributions. Together, these two signals imply that the object has jumped $r + c = 15^\circ - 10^\circ = 5^\circ$. But the prior probability of a jump, j , (black distribution marked j) is small away from zero, so the percept, j^* , is 'pulled' towards zero. The degree of pull depends on the widths of the c , r and j distributions (Fig. 3b–d).

Simulations indicate that perceived jumps should be a nonlinear function of actual jump size (Fig. 3e). Around zero, the graph runs almost horizontally, indicating strong contraction. Also, the graph is shifted in the direction of the saccade, because immediately after a saccade, the eye-position signal c is hypometric²². We tested these predictions experimentally. Eight subjects made saccades towards a target that jumped parallel to the saccade. Using a mouse, they then moved the target back to what they felt was its presaccadic position, thereby indicating the perceived jump size. As in the simulations, all subjects showed a contracted and shifted percept (Fig. 3f).

Studies have revealed various patterns of compression and shift in perisaccadic spatial percepts, similar but not identical to those in Fig. 3f, for paradigms more complex than our continuous presentation of a single visible target^{1–6}. These tasks are beyond the current scope of our model, because some involve multiple visible objects, which complicate the probabilities, and all use a flashed target, whose perisaccadic localization requires precise judgements of timing and which the brain may interpret as a new, second object rather than as a reappearance of the presaccadic target, again affecting its probable location. For simplicity, the model considers just one object and two time points (pre- and postsaccade), but the

principle of optimal integration could be extended to deal with more complex conditions.

Our model of transsaccadic integration reconciles and clarifies previous approaches. It shows that the brain uses eye-position signals in interpreting retinal information¹⁸, but only to a limited extent²⁷. Our simulations show that SSD arises from noisy sensorimotor signals^{7,23} and from the prior implausibility of sudden jumps during saccades²⁸. And we show that optimal integration can explain both SSD and graded distortions of spatial perception.

These findings support a view in which the brain pieces together a coherent percept of the world like a probabilistic, spatiotemporal jigsaw puzzle. The computations are complex, but they could be learned from available sensory and motor signals (they are well approximated by artificial networks of just a few hundred sigmoid neurons). We suggest that natural selection and learning shape the brain's circuits into transsaccadic integrators that make optimal use of each piece of sensorimotor information. □

Methods

Optimal transsaccadic integration

For a full account of the optimized visuomotor model and its derivation, see Supplementary Information. The system receives three input variables: r is the change in the target's retinal location, before and after the saccade; v is an independent estimate of that change in retinal location derived from visual velocity detectors; c is the estimated change in eye position, conveyed by efference copy or proprioception. These signals are influenced by two other variables, s and j , which represent saccades and jumps; that is, the amplitudes and directions of sudden movements of the visible object. From c , r and v , the system guesses j . Using Bayes' rule and other laws of probability, it can be shown that the optimal estimate, j^* (the one that will yield the smallest misjudgement of j on average), is the ratio of two iterated integrals:

$$j^* = \frac{\int ds p(s) p(j) p(c|s) p(r|js) p(v|js)}{\int dj \int ds p(s) p(j) p(c|s) p(r|js) p(v|js)} \quad (1)$$

Here, each p is a probability density function; for example, $p(r|js)$ is the conditional probability density that the retinal shift would have value r if the jump had value j and the saccade had value s .

All distributions in the model are realistic, insofar as the real parameters are known, although at present, $p(j)$ and $p(r|js)$ can only be estimated. $p(s)$ is gaussian, with s.d. $\sigma_s = 0.01d + 0.05^\circ$ (where d is the target distance) orthogonal to the saccade²⁹; to simulate parallel scatter, we multiplied this value by the scatter ratios 2.5, 3, 3.5, 4 and 4.5 (compare Fig. 2a). Mean saccade size μ_s is equal to the target distance minus an undershoot, u , which is well approximated in our data by $u = 0.23\sigma_s + 0.41^\circ$. $p(j)$ is a Laplace distribution, $\exp(-|j|/w)/2w$, where w is a free parameter that we set to 0.065° normally, and to 0.13° in the presence of blanking in Fig. 1f. $p(c|s)$, $p(r|js)$ and $p(v|js)$ are all gaussian. Because c is delayed and therefore hypometric right after saccades²², $\mu_c = Ds$ (where D , the factor by which saccade size is underestimated, varies between about 0.75 and 0.9); we set $\sigma_c = \sigma_s$. For r , we set $\mu_r = j - s$ and $\sigma_r = 0.02e + 0.15^\circ$, where e is the postsaccadic, or postjump, eccentricity of the target (initial eccentricity was constant for any one experiment). And for v , we set $\mu_v = r$ when there was no saccade, and $\mu_v = 0$ otherwise; we set σ_v at 0.005° (ref. 12). To simulate neural variability in the integrator itself, we added gaussian noise of s.d. 0.004 radians to j^* . Moderate changes in these parameters do not affect the qualitative predictions—such as SSD, shifts and compression—and the linear relationships in Figs 1a and 2.

Experiments

Seven subjects with normal or corrected-to-normal vision participated in the first experiment. They sat in a dark room with their head stabilized on a bite-bar, and viewed computer-generated stimuli (20 cd m^{-2}) back-projected onto a 1.4-by-1.9-m screen spanning about 100° horizontally and 90° vertically. At the start of each trial, the subject fixated a small dot, 0.4° across, at one of four possible locations, at the corners of an imaginary central 10° -by- 10° square (rather than straight ahead, so subjects could not tell the direction of the target's subsequent jump by its final location relative to the midline of the head). After 0.5–1 s, the fixation dot vanished and a target spot, 0.8° across, appeared 15° to the left, right, above or below fixation. Subjects looked to the target as quickly as possible. The computer, monitoring eye position from search-coil signals at 1,000 Hz, detected the saccade (eye velocity, $\geq 36^\circ \text{ s}^{-1}$; eye position, $\geq 1.5^\circ$ from the fixation point) and moved the target horizontally or vertically; 200 ms later, the screen went white and the subjects chose the direction in which they believed the target had jumped. An adaptive algorithm³⁰ varied the jump from trial to trial, mapping out the thresholds of perception. Thresholds differed for onward versus backward jumps, as predicted by the model; Fig. 2 is based on thresholds for pooled onward and backward jumps, but data and model agreed equally well when we separated these jump directions.

We switched off the fixation dot so that subjects could not use its direction relative to the target to perceive orthogonal jumps. Others have found reduced SSD orthogonal to

saccades, although their effects were smaller than ours, probably for other, methodological reasons⁷.

In Fig. 2b, we quantified our subjects' uncertainty about eye and target position using scatter ratios. We validated this method with a study identical to the first, except that the target vanished at the start of the saccade. Subjects made as many corrective saccades as they wished, then pressed a button when they felt they were looking at the former location of the target. Clearly, the scatter of eye positions at this moment reflected the subject's uncertainty about the placement of eye and target. But the scatter ratios at that moment correlated tightly ($r = 0.93$, $P < 0.05$) with scatter ratios after the initial saccade, before any corrective movements. This shows that we can take, as an accurate measure of uncertainty, the scatter ratio after initial, uncorrected saccades, as we did in Fig. 2b.

The experiments on graded contraction of jumps were again identical to the first study, except that eight subjects made horizontal saccades to a target that jumped, during the saccade, to a location chosen randomly from a horizontal range of $\pm 10^\circ$; 200 ms after the saccade, the target was replaced by a vertical bar that the subject moved to what they felt was the target's presaccadic location. Only one object was visible at any one time. In other experiments, subjects moved the target itself rather than a bar, but the results were the same.

Received 22 November 2002; accepted 20 January 2003; doi:10.1038/nature01439.

1. Matin, L. & Pearce, D. G. Visual perception of direction for stimuli flashed during voluntary saccadic eye movements. *Science* **148**, 1485–1488 (1965).
2. Honda, H. Saccade-contingent displacement of the apparent position of visual stimuli flashed on a dimly illuminated structured background. *Vision Res.* **33**, 709–716 (1993).
3. Ross, J., Morrone, M. C. & Burr, D. C. Compression of space before saccades. *Nature* **386**, 598–601 (1997).
4. Morrone, M. C., Ross, R. & Burr, D. C. Apparent position of visual targets during real and simulated saccadic eye movements. *J. Neurosci.* **17**, 7941–7953 (1997).
5. Cai, R. H., Pouget, A., Schlag-Rey, M. & Schlag, J. Perceived geometrical relationships affected by eye-movement signals. *Nature* **386**, 601–604 (1997).
6. Lappe, M., Awatier, H. & Krekelberg, B. Postsaccadic visual references generate presaccadic compression of space. *Nature* **403**, 892–895 (2000).
7. Mack, A. An investigation of the relationship between eye and retinal image movement in the perception of movement. *Percept. Psychophys.* **8**, 291–298 (1970).
8. Bridgeman, B., Hendry, D. & Stark, L. Failure to detect displacement of the visual world during saccadic eye movements. *Vision Res.* **15**, 719–722 (1975).
9. Westheimer, G. Spatial vision. *Annu. Rev. Psychol.* **35**, 201–226 (1984).
10. Carpenter, R. H. S. in *Vision and Visual Dysfunctions* (ed. Carpenter, R. H. S.) 1–10 (Macmillan, Boca Raton, 1991).
11. Choudhury, B. P. & Crossey, A. D. Slow-movement sensitivity in the human field of vision. *Physiol. Behav.* **26**, 125–128 (1981).
12. Tynan, P. D. & Sekuler, R. Motion processing in peripheral vision: Reaction time and perceived velocity. *Vision Res.* **22**, 61–68 (1982).
13. Burr, D. C., Holt, J., Johnstone, J. R. & Ross, J. Selective depression of motion sensitivity during saccades. *J. Physiol. (Lond.)* **333**, 1–15 (1982).
14. Burr, D. C., Morrone, M. C. & Ross, J. Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature* **371**, 511–513 (1994).
15. Shiori, S. & Cavanagh, P. Saccadic suppression of low-level motion. *Vision Res.* **29**, 915–928 (1989).
16. Ilg, U. & Hoffmann, K.-P. Motion perception during saccades. *Vision Res.* **33**, 211–220 (1993).
17. Steinbach, M. J. Proprioceptive knowledge of eye position. *Vision Res.* **27**, 1737–1744 (1986).
18. von Helmholtz, H. *Handbuch der physiologischen Optik* (Leopold Voss, Hamburg and Leipzig, 1910).
19. von Holst, E. & Mittelstaedt, H. Das Refferenzprinzip (Zwischenwirkung zwischen Zentralnervensystem und Peripherie). *Naturwissenschaften* **37**, 464–476 (1950).
20. Sperry, R. W. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.* **43**, 482–489 (1950).
21. Dassonville, P., Schlag, J. & Schlag-Rey, M. Oculomotor localization relies on a dampened representation of saccadic eye movement displacement in human and nonhuman primates. *Vis. Neurosci.* **9**, 261–269 (1992).
22. Honda, H. Interaction of extraretinal eye position signals in a double-step saccade task: Psychophysical estimation. *Exp. Brain Res.* **113**, 327–336 (1997).
23. Li, W. & Matin, L. The influence of saccade length on the saccadic suppression of displacement detection. *Percept. Psychophys.* **48**, 453–458 (1990).
24. Li, W. & Matin, L. Saccadic suppression of displacement: Separate influences of saccade size and of target retinal eccentricity. *Vision Res.* **37**, 1779–1797 (1997).
25. Gysen, V., De Graef, P. & Verfaillie, K. Detection of intrasaccadic displacements and depth rotations of moving objects. *Vision Res.* **42**, 379–391 (2002).
26. Deubel, H., Schneider, W. X. & Bridgeman, B. Postsaccadic target blanking prevents saccadic suppression of image displacement. *Vision Res.* **36**, 985–996 (1996).
27. Bridgeman, B., Van der Heijden, A. H. C. & Velichovsky, B. M. A theory of visual stability across saccadic eye movements. *Behav. Brain Sci.* **17**, 247–292 (1994).
28. MacKay, D. M. in *Handbook of Sensory Physiology* Vol. VII/3 (ed. Jung, R.) 307–332 (Springer, Berlin, 1973).
29. Van Opstal, A. J. & Van Gisbergen, J. A. M. Scatter in the metrics of saccades and properties of the collicular motor map. *Vision Res.* **29**, 1183–1196 (1989).
30. Kontsevich, L. L. & Tyler, C. W. Bayesian adaptive estimation of psychometric slope and threshold. *Vision Res.* **39**, 2729–2737 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank S. Sun for invaluable programming, H. Wang for help with experiments, and I. Frenkel, P. Hallett, C. Hawkins, D. Henriques, E. Klier, J. C. Martinez Trujillo, P. Medendorp and K. Schreiber for comments on the manuscript. This study was supported by the Canadian Institutes for Health Research. J.D.C. is supported by the Canadian Research Chair Program.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.T. (e-mail: douglas.tweed@utoronto.ca).

Monoclonal antibodies inhibit prion replication and delay the development of prion disease

Anthony R. White*, Perry Enever*, Mourad Tayebi*, Rosey Mushens†, Jackie Linehan‡, Sebastian Brandner‡, David Anstee†, John Collinge*‡ & Simon Hawke*

* CNS Infection and Immunity Group, Department of Neurogenetics, Division of Neurosciences and Psychological Medicine, Faculty of Medicine, Imperial College, Norfolk Place, London W2 1PG, UK

† IBGRL, National Blood Service, Bristol BS10 5ND, UK and ‡ MRC Prion Unit and Department of Neurodegenerative Disease, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK

Prion diseases such as Creutzfeldt–Jakob disease (CJD) are fatal, neuro-degenerative disorders with no known therapy. A proportion of the UK population has been exposed to a bovine spongiform encephalopathy-like prion strain^{1–3} and are at risk of developing variant CJD⁴. A hallmark of prion disease is the transformation of normal cellular prion protein (PrP^C) into an infectious disease-associated isoform⁵, PrP^{Sc}. Recent *in vitro* studies indicate that anti-PrP monoclonal antibodies with little or no affinity for PrP^{Sc} can prevent the incorporation of PrP^C into propagating prions^{6,7}. We therefore investigated in a murine scrapie model whether anti-PrP monoclonal antibodies show similar inhibitory effects on prion replication *in vivo*. We found that peripheral PrP^{Sc} levels and prion infectivity were markedly reduced, even when the antibodies were first administered at the point of near maximal accumulation of PrP^{Sc} in the spleen. Furthermore, animals in which the treatment was continued remained healthy for over 300 days after equivalent untreated animals had succumbed to the disease. These findings indicate that immunotherapeutic strategies for human prion diseases are worth pursuing.

Recombinant human PrP^{91–231} folded into either α - or β -conformations^{8,9} was used to produce monoclonal antibodies in mice lacking PrP (*Prnp*^{0/0})¹⁰ that are intolerant to PrP^C. ICSM 35, an immunoglobulin- γ 2b (IgG2b) monoclonal antibody raised against β -PrP, with high affinity for both murine PrP^C and PrP^{Sc} (Fig. 1a; A. Khalili-Shirazi, S.H. and J.C., unpublished data), recognizes a region between amino acid residues 91 and 110 (ref. 11). ICSM 18 (isotype IgG1), raised against α -PrP, recognizes residues 146–159 of murine PrP and has a lower affinity for PrP^{Sc} (Fig. 1a). FVB/N mice were challenged intraperitoneally (i.p.) with Rocky Mountain Laboratory (RML) scrapie brain homogenate derived from terminally scrapie-sick mice and treated with ICSM 35, ICSM 18 or isotype control antibodies BRIC 126 (IgG2b) and BRIC 222 (IgG1) by twice weekly i.p. injection (2 mg per injection) from 7 or 30 days post inoculation (p.i.). Enzyme-linked immunosorbent assay (ELISA) analysis after 30 days of antibody treatment revealed no

significant differences between ICSM 35 or ICSM 18 antibody levels in the serum (Supplementary Information). Western blots of proteinase K-treated, phosphotungstic-acid-precipitated PrP^{Sc} from spleens of mice at 60 days p.i. revealed that treatment from 7 days p.i. with ICSM 35 or ICSM 18, but not with BRIC antibodies,

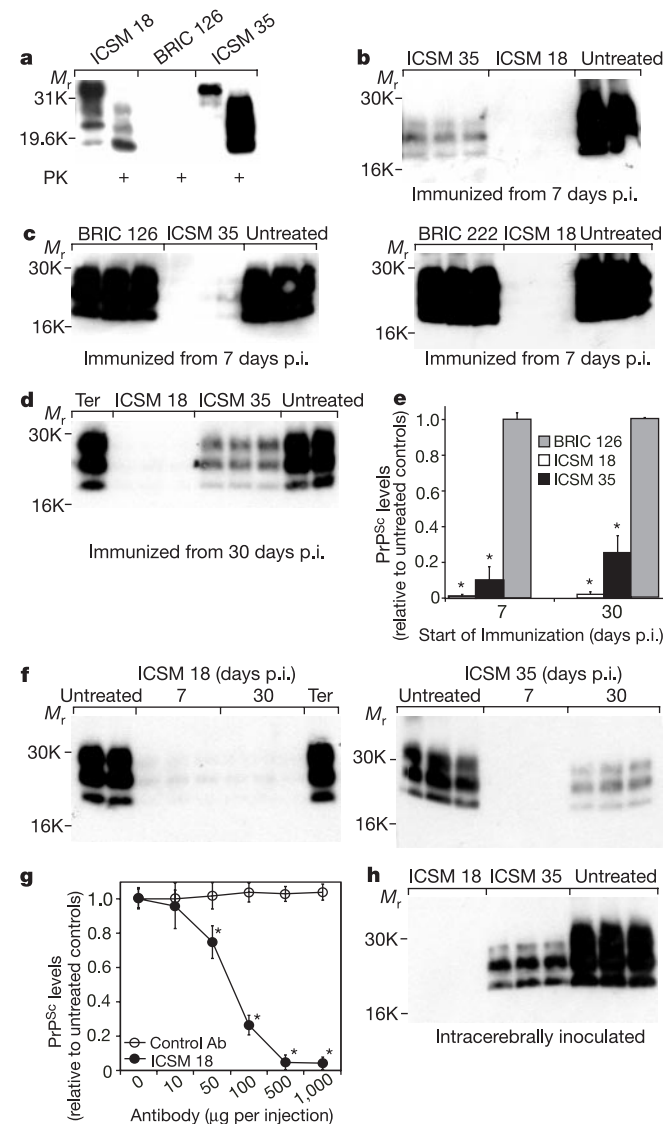


Figure 1 Anti-PrP antibodies inhibit splenic PrP^{Sc} protein levels. Western blots of proteinase K-digested, phosphotungstic-acid-precipitated PrP^{Sc} from spleens of mice 60 days p.i. with RML scrapie. Mice were inoculated i.p. except in **h**. Each lane contains PrP^{Sc} from an individual mouse. *M_r*, relative molecular mass; Ter, pooled splenic PrP^{Sc} from mice succumbing to terminal scrapie (195 ± 5 days p.i.). **a**, Immunoprecipitation of PrP from scrapie-infected mouse brain using ICSM or BRIC antibodies. PK, proteinase K. **b**, ICSM 18 and ICSM 35 induced substantial reductions in splenic PrP^{Sc} levels when treatment began from 7 days p.i., but this reduction was not seen in **c**, the BRIC-control-treated mice. **d**, ICSM 18 and ICSM 35 reduced splenic PrP^{Sc} when treatment began at 30 days p.i. **e**, Densitometry of PrP^{Sc} levels in the western blots. ICSM 18 induced greater reduction in PrP^{Sc} levels in spleens than ICSM 35, whereas BRIC 126 had no effect (asterisk, *P* < 0.001 compared with untreated spleens). **f**, ICSM 18 induced efficient clearance of PrP^{Sc} whether treatment began at 7 or 30 days p.i. ICSM 35 induced more efficient inhibition of PrP^{Sc} accumulation when treatment began at 7 days rather than 30 days p.i. **g**, ICSM 18 induced a dose-dependent reduction in PrP^{Sc} levels in spleens as determined by densitometry of western blots (asterisk, *P* < 0.001, ANOVA compared with control antibody (Ab) treatment). **h**, ICSM 18 and ICSM 35 inhibited splenic PrP^{Sc} accumulation in i.c. inoculated mice.

markedly inhibited PrP^{Sc} accumulation in the spleen (Fig. 1b, c).

Time course analysis of peripheral PrP^{Sc} accumulation in mice confirmed that PrP^{Sc} was detectable in the spleen at 7 days and plateaued by 30–40 days after peripheral challenge (data not shown) as previously reported¹². Treatment of scrapie-infected mice with ICSM 35 or ICSM 18 from 30 to 60 days p.i. resulted in a substantial reduction in splenic PrP^{Sc} levels when compared with untreated controls (Fig. 1d) or mice treated with isotype control antibody (Fig. 1e). ICSM 18 reduced PrP^{Sc} levels by $99 \pm 1\%$ and $96 \pm 3\%$ (mean $\pm$ s.d., $P < 0.001$, analysis of variance (ANOVA)) after treatment from 7 or 30 days p.i., respectively (Fig. 1e). ICSM 35 also lowered PrP^{Sc} levels by $90 \pm 8\%$ and $75 \pm 3\%$ ($P < 0.001$, ANOVA) for the same treatment periods, whereas isotype control antibody did not alter PrP^{Sc} levels (Fig. 1e). Immunoblots probed with ICSM 35 instead of ICSM 18 as the primary antibody produced indistinguishable results (data not shown).

Treatment of mice from either 7 or 30 days p.i. with ICSM 18 consistently resulted in almost complete loss of detectable PrP^{Sc} in spleens by western blot (Fig. 1f). However, although ICSM 35 treatment from 7 days p.i. reduced splenic PrP^{Sc} levels with similar efficiency to ICSM 18, treatment with ICSM 35 from 30 days p.i. led to less complete inhibition of PrP^{Sc} replication (Fig. 1f). This may reflect differences in affinity or avidity of ICSM 35 and ICSM 18 for normal and disease-related PrP, as shown in Fig. 1a. Reduction in splenic PrP^{Sc} levels induced by ICSM 18 was dose-dependent (Fig. 1g). Analogous results were obtained when intracerebrally inoculated (i.c.) mice were treated with ICSM 18 or ICSM 35 (Fig. 1h). As ICSM 18 recognizes the PrP epitope at residues 146–159, these data are consistent with residues 132–156—which incorporate helix 1—of mouse PrP having a crucial role in prion replication^{6–7,13–15}.

Tissue sections derived from formalin-fixed spleens at 60 days p.i. were examined by standard PrP immunohistochemistry using ICSM 18 and ICSM 35 (Fig. 2a–d). Both monoclonal antibodies were applied to adjacent splenic sections to ensure that antibody bound *in vivo* did not spuriously block the detecting antibody. ICSM 35 revealed intense labelling of PrP^{Sc}, primarily associated with germinal centres (Fig. 2a). PrP^{Sc} immunostaining was substantially reduced in spleens of ICSM 18- and ICSM 35-treated mice when labelled with ICSM 35 as primary antibody (Fig. 2b, c; see also Supplementary Information). ICSM 18 produced only weak staining of PrP^{Sc} in spleens from untreated, scrapie-infected mice and this staining was also diminished in both ICSM 35- and ICSM 18-treated mice (data not shown). As expected, PrP^{Sc}-positive germinal centres were not seen in splenic sections from mice that had not been challenged with RML prions (Fig. 2d). Bioassay of splenic homogenates from ICSM 35- and ICSM 18-treated mice showed a 1.5–3.5 and greater than 4 log reduction in infectious titres, respectively, compared with controls (Table 1).

We then sought evidence for accompanying benefit to the health of the mice. Although treatment was ineffective if begun at the onset of clinical scrapie or after intracerebral challenge (Table 1), mice that were inoculated i.p. with RML scrapie and treated with ICSM 35 or ICSM 18 (2 mg twice weekly) from 7 and 30 days p.i. survived for much longer than untreated mice or mice treated with isotype control antibody (Table 1). The mean survival for untreated i.p.-inoculated mice was 197 ± 5 days. Mice treated with ICSM 18 or ICSM 35 from 7 or 30 days post i.p. challenge survived for more than 500 days p.i. (Table 1); an extension of the incubation period by at least 153% ($P < 0.001$, ANOVA). No clinical signs of scrapie (as described in the Methods) or weight loss (Supplementary Fig. 1) have yet been observed in these mice. Experiments are in progress to determine whether prion infection has been suppressed or eradicated in these animals and if the latter, what period of treatment is required to effect a cure.

Tissues were examined from mice lacking clinical signs of scrapie after treatment with ICSM 18 or ICSM 35 from 7 days p.i. In these

mice (killed at 250 days p.i.), PrP^{Sc} was undetectable in the brain after either of the two treatments, and only low levels of PrP^{Sc} were seen in spleens of mice treated with ICSM 35 (Fig. 3a–d). Similarly, no PrP^{Sc} was observed in the brain after ICSM 35 treatment from 30 days p.i. and analysed at 230 days p.i. (data not shown). High levels of PrP^{Sc} were observed in brains and spleens from BRIC antibody-treated and untreated mice that succumbed to scrapie after i.p. inoculation (Fig. 3a–d). These findings were confirmed by histopathological analysis (data not shown), and bioassay of infectivity from these tissues is in progress.

We next examined whether passive immunization with anti-PrP monoclonal antibodies affected immune cell populations, as targeted depletion of PrP⁺ cell types could have contributed to loss of PrP^{Sc} detection in spleens of antibody-treated mice^{16,17}. No differences were observed in splenic T- and B-cell populations between

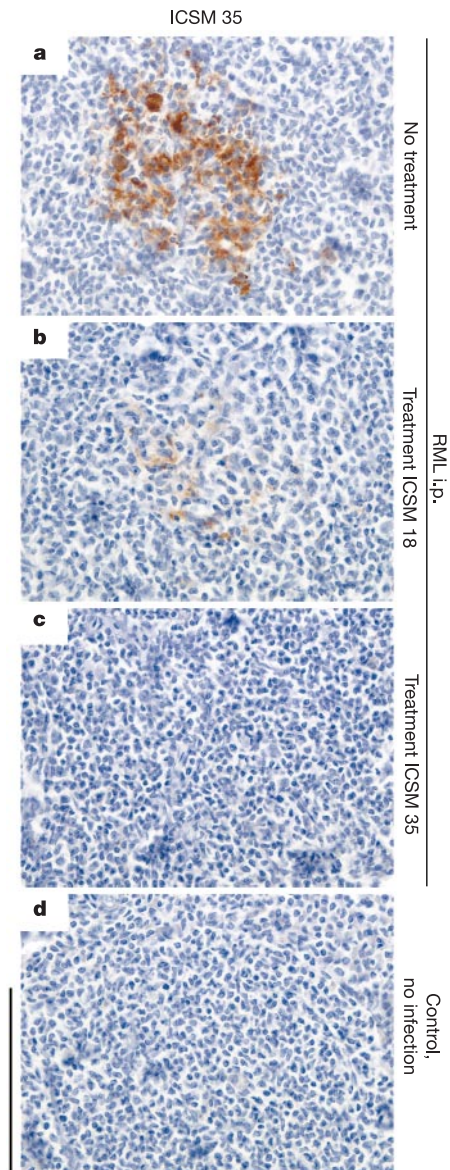


Figure 2 Immunohistochemistry of spleens from antibody-treated mice.

a–d, Immunohistochemical staining of spleens from ICSM 35-treated mice 60 days after i.p. inoculation with RML scrapie: **a**, immunostaining with ICSM 35 reveals strong PrP immunoreactivity (red) in follicle centres in spleens of scrapie-infected mice. **b**, After treatment with ICSM 18, very few follicles are immunoreactive for PrP. **c**, No positive follicles detectable after treatment with ICSM 35. **d**, No positive follicles are detectable in control spleens. Scale bar, 100 μ m.

Table 1 Effect of passive immunization with ICSM antibodies on spleen infectivity and survival of FVB/N mice inoculated with RML scrapie

i.c. or i.p. inoculation*	Antibody treatment†	Start of treatment (days p.i.)	Spleen bioassay§		FVB/N mice succumbing to scrapie (n/n ₀)	Mean survival time (days p.i.)
			Mortality	Spleen titre		
i.c.	ICSM 35	7	-	-	6/6	152 ± 7
i.c.	ICSM 35	CO‡	-	-	8/8	148 ± 10
i.c.	ICSM 18	7	-	-	5/5	151 ± 2
i.c.	ICSM 18	CO	-	-	7/7	149 ± 3
i.c.	BRIC 126	7	-	-	5/5	147 ± 4
i.c.	None	-	-	-	11/11	151 ± 7
i.p.	ICSM 35	7	15/15	3.5 (2.7–4.3)	0/6	>500
i.p.	ICSM 35	30	-	-	0/5	>500
i.p.	ICSM 35	CO	-	-	6/6	193 ± 4
i.p.	ICSM 18	7	3/12	<1.5	0/6¶	>500
i.p.	ICSM 18	30	1/10	<1.5	0/5	>500
i.p.	ICSM 18	CO	-	-	5/5	193 ± 5
i.p.	BRIC 126	7	9/9	6.0	6/6	195 ± 7
i.p.	BRIC 126	30	-	-	4/4	198 ± 4
i.p.	None	-	13/13	6.0	15/15	197 ± 5

*The infectious titre of the pooled RML scrapie brain homogenate was determined as 8.1 log LD₅₀ per g brain by infectivity assay with *tga20* indicator mice²⁶. FVB/N mice were inoculated intracerebrally (i.c.) with 30 µl or intraperitoneally (i.p.) with 100 µl of 1% homogenate.

†Twice weekly until 375 days p.i., then once per week, 2 mg per injection.

‡CO, treatment began at clinical onset (see Methods); 129–136 days p.i. (i.c. inoculated mice) and 168–177 days p.i. (i.p. inoculated mice).

§Determined at 60 days p.i. of passively immunized and untreated FVB/N mice. Mortality, *tga20* mice succumbing to scrapie (n/n₀), where n/n₀ is the number of animals succumbing to scrapie per number of animals inoculated; spleen titre, mean log LD₅₀ infectious units per ml 10% homogenate.

||Range of log LD₅₀ infectious units per ml 10% homogenate.

¶Two mice died of unrelated deaths at 450 and 480 days p.i.

untreated scrapie-inoculated mice and antibody-treated mice (examined at 60 days p.i.) as determined by immunostaining of cryostat tissue sections or by flow cytometry (see Supplementary Fig. 2 and Supplementary Table 1). Follicular dendritic cells (FDCs) are an important site of PrP^{Sc} accumulation after peripheral prion infection^{17,18}. Immunostaining of splenic sections from untreated mice revealed FDC-M1⁺ cells scattered throughout the germinal centres or in small clusters within the germinal centre (Supplementary Fig. 2). By contrast, ICSM or BRIC antibody-treated spleens often revealed larger numbers of FDC-M1⁺ cells clustered tightly within the germinal centre (Supplementary Fig. 2). These data clearly show that prolonged treatment with the anti-PrP antibodies does not induce depletion of the FDC cell type within the spleen, and strongly suggest that the substantial reduction in PrP^{Sc} levels observed in spleens treated with anti-PrP antibodies is probably due to direct inhibition of PrP^{Sc} production. This was further supported by unaltered PrP^C protein levels in spleen homogenates from mice treated with ICSM 35 or ICSM 18 (Supplementary Fig. 3).

Our studies demonstrate that substantial peripheral prion replication can be effectively suppressed by passive immunization. Importantly, treatment began well after the onset of peripheral prion replication and in the case of the 30 days p.i. treatment group, during the plateau phase of PrP^{Sc} accumulation¹². Continued treatment delayed onset of scrapie by more than 150% of the usual incubation period in wild type FVB/N mice. In contrast, almost all previous therapeutic interventions only show benefit if treatment is begun before, or very soon after the day of inoculation^{19–24}, probably reflecting simple neutralization of the inoculum.

It is important to emphasize that passive transfer of these anti-PrP antibodies had no effect late in the incubation period when clinical signs had developed, probably reflecting inadequate translocation of anti-PrP antibody across the blood-brain barrier²⁵. Furthermore, although we found no evidence for autoimmune reactions in mice, when treating humans infected with CJD or other prion diseases with humanized forms of these and other anti-PrP monoclonal antibodies, unwanted autoimmune side effects could develop. Despite these caveats, our findings are encouraging and provide an important impetus for pursuing prion immunotherapeutics. □

Methods

Production of monoclonal antibodies

ICSM 35 and ICSM 18 monoclonal antibodies were produced as described¹¹. ICSM and BRIC monoclonal antibodies were identically affinity purified from culture supernatant, concentrated, and stored as sterile solutions without vehicle protein at 4 °C. They were used undiluted or diluted in PBS before use *in vivo*.

Inoculation of FVB/N mice with RML prion inoculum

RML prions were passaged in FVB/N mice and prion inoculum was prepared from the brains of terminally sick mice (incubation time to terminal scrapie, 153 ± 4 days). Brains were homogenized in PBS (10% w/v) with 1% bovine serum albumin (BSA) using a Ribolyzer (Hybaid). The homogenate was spun for 5 min at 500g and supernatants pooled and frozen at –80 °C until use. The infectious titre of the pooled homogenate was determined as 8.1 log LD₅₀ (lethal dose to 50% of animals) per g brain by infectivity assay with *tga20* indicator mice²⁶. FVB/N mice were inoculated i.c. with 30 µl or i.p. with 100 µl of 1% homogenate.

Infectivity bioassays

Assays were performed on 1% spleen homogenates. Aliquots of 30 µl were inoculated i.c. into groups of three or four *tga20* mice per treatment regimen. Incubation time to terminal scrapie sickness was determined and infectivity titres were calculated by using the equation $y = 11.45 - 0.088x$ (for RML 4.1), where y is the infectious titre (log LD₅₀), and x is the incubation time (in days) to terminal disease²⁶.

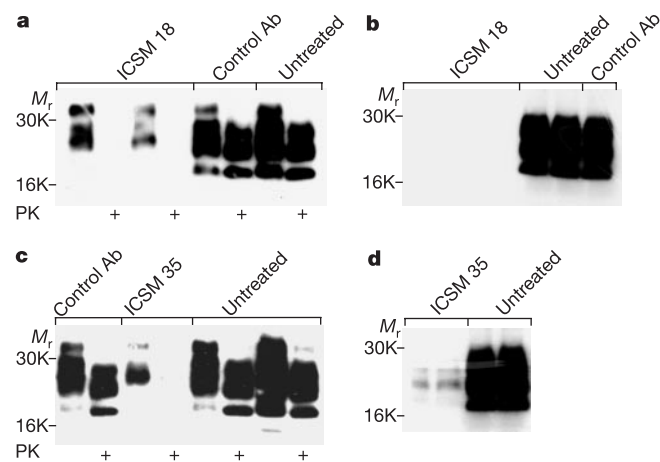


Figure 3 PrP^{Sc} protein levels in mice with extended survival. Western blots of PrP^{Sc} in brain homogenates (**a, c**) or proteinase K (PK)-treated, phosphotungstic-acid-precipitated PrP from spleens (**b, d**). PrP^{Sc} was undetectable (**a–c**) or significantly reduced (**d**) in the brains and spleens of mice treated with ICSM 18 or ICSM 35 from 7 days p.i. and killed 250 days p.i. Mice treated with BRIC control antibody (Ab) succumbed to scrapie after 195 days p.i.

Passive immunization

Groups of mice were injected twice weekly through the intraperitoneal route with 2 mg (unless otherwise stated) ICSM 18, ICSM 35, IgG1 isotype control (BRIC 222 recognizing CD44; ref. 27) or IgG2b isotype control (BRIC 126 recognizing CD47; ref. 28) antibodies in PBS. Animals were monitored daily for clinical symptoms of scrapie²⁹ and weighed weekly from 17 weeks after i.c. inoculation or 22 weeks after i.p. inoculation. Clinical signs in untreated mice were first observed approximately 4 weeks before terminal illness (day of death) and included coat ruffling/dyscoloration, progressive weight loss, bradykinesia (slow movement), tail rigidity, dystonia (clasp foot), kyphosis (hunched back), ataxia and stupor. Weights of scrapie-infected (untreated) mice decreased before terminal illness from 3 and 4 weeks in i.c.- and i.p.-inoculated mice, respectively (Supplementary Fig. 1). Confirmation of scrapie disease was performed by western blot analysis of PrP^{Sc} in brain tissue and in some cases by standard PrP immunohistochemistry.

Immunoprecipitation

Immunoprecipitation of PrP from murine brain tissues using ICSM and BRIC antibodies was performed as described¹⁴.

Western-blot analysis

Spleens were homogenized in PBS to 10% w/v and PrP^{Sc} was precipitated from 500 µl of homogenates using sodium phosphotungstic acid (NaPTA) as previously described³⁰. PrP^{Sc} pellets were resuspended in 20 µl of 2% sarkosyl buffer, treated with proteinase K (50 µg ml⁻¹, 50 min, 37 °C), boiled in sample buffer (5 min) and 15 µl aliquots (equivalent to about 2 mg spleen homogenate) were electrophoresed through 16% SDS–polyacrylamide electrophoresis gels. Brain homogenates were diluted to 1% (w/v) in PBS, treated with proteinase K (50 µg ml⁻¹ for 60 min, 37 °C) and electrophoresed as for spleens. Proteins were transferred to polyvinylidene fluoride (PVDF) membrane by semi-dry blotting, blocked with TBST/5% non-fat milk, incubated with biotinylated ICSM 18 (0.1 µg ml⁻¹) and developed by enhanced chemiluminescence (Amersham). Semi-quantification was performed by densitometric analysis using MacBas version 2.5 software. At least three to four mice were examined from each treatment group. Bars on graphs are standard deviations. To standardize the PrP^{Sc} signal between blots, 10 µl of PrP^{Sc} precipitated from pooled spleens of terminal (Ter) scrapie-affected mice was loaded on each gel. Densitometric measurement of PrP^{Sc} from treated and untreated spleens was compared to the standard PrP^{Sc} sample and adjusted to relative intensities.

Histology and immunohistochemistry

For PrP^{Sc} immunohistochemistry, spleens and brains were fixed in 10% formalin. Prion infectivity was inactivated by immersion in 98% formic acid and postfixed in formalin for 24 h. Tissues were dehydrated in graded alcohols and xylene, embedded in paraffin, sections cut at 3 µm nominal thickness and stained with haematoxylin and eosin. After antigen retrieval by microwave treatment for 15 min, sections were immunostained with biotinylated ICSM 18 or ICSM 35 on a Ventana automated staining apparatus.

Received 20 September 2002; accepted 27 January 2003; doi:10.1038/nature01457.

- Collinge, J., Sidle, K. C., Meads, J., Ironside, J. & Hill, A. F. Molecular analysis of prion strain variation and the aetiology of 'new variant' CJD. *Nature* **383**, 685–690 (1996).
- Bruce, M. E. *et al.* Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent. *Nature* **389**, 498–501 (1997).
- Hill, A. F. *et al.* The same prion strain causes vCJD and BSE. *Nature* **389**, 448–450 (1997).
- Will, R. G. *et al.* A new variant of Creutzfeldt-Jakob disease in the UK. *Lancet* **347**, 921–925 (1996).
- Prusiner, S. B. Prions. *Proc. Natl. Acad. Sci. USA* **95**, 13363–13383 (1998).
- Enari, M., Flechsig, E. & Weissmann, C. Scrapie prion protein accumulation by scrapie-infected neuroblastoma cells abrogated by exposure to a prion protein antibody. *Proc. Natl. Acad. Sci. USA* **98**, 9295–9299 (2001).
- Peretz, D. *et al.* Antibodies inhibit prion propagation and clear cell cultures of prion infectivity. *Nature* **412**, 739–743 (2001).
- Jackson, G. S. *et al.* Reversible conversion of monomeric human prion protein between native and fibrillogenic conformations. *Science* **283**, 1935–1937 (1999).
- Jackson, G. S. *et al.* Multiple folding pathways for heterologously expressed human prion protein. *Biochim. Biophys. Acta* **1431**, 1–13 (1999).
- Bueler, H. *et al.* Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein. *Nature* **356**, 577–582 (1992).
- Beringue, V. *et al.* Regional heterogeneity of cellular prion protein isoforms in the brain. *Brain* (submitted).
- Beringue, V. *et al.* Opposite effects of dextran sulfate 500, the polyene antibiotic MS-8209, and Congo red on accumulation of the protease-resistant isoform of PrP in the spleens of mice inoculated intraperitoneally with the scrapie agent. *J. Virol.* **74**, 5432–5440 (2000).
- Souan, L. *et al.* Modulation of proteinase-K resistant prion protein by prion peptide immunisation. *Eur. J. Immunol.* **31**, 2338–2346 (2001).
- Heppner, F. L. *et al.* Prevention of scrapie pathogenesis by transgenic expression of anti-prion protein antibodies. *Science* **294**, 178–182 (2001).
- Westaway, D. & Carlson, G. A. Mammalian prion proteins: enigma, variation and vaccination. *Trends Biochem. Sci.* **27**, 301–307 (2002).
- Klein, M. A. *et al.* PrP expression in B lymphocytes is not required for prion neuroinvasion. *Nature Med.* **4**, 1429–1433 (1998).
- Montrasio, F. *et al.* Impaired prion replication in spleens of mice lacking functional follicular dendritic cells. *Science* **288**, 1257–1259 (2000).
- Brown, K. L. *et al.* Scrapie replication in lymphoid tissues depends on prion protein-expressing follicular dendritic cells. *Nature Med.* **5**, 1308–1312 (1999).
- McKenzie, D., Kaczowski, J., Marsh, R. & Aiken, J. Amphotericin B delays both scrapie agent replication and PrP-res accumulation early in infection. *J. Virol.* **68**, 7534–7536 (1994).
- Ingrassio, L., Ladogana, A. & Pocchiari, M. Congo red prolongs the incubation period in scrapie-

infected hamsters. *J. Virol.* **69**, 506–508 (1995).

- Farquhar, C., Dickinson, A. & Bruce, M. Prophylactic potential of pentosan polysulphate in transmissible spongiform encephalopathies. *Lancet* **353**, 117 (1999).
- Brown, P. Drug therapy in human and experimental transmissible spongiform encephalopathy. *Neurology* **58**, 1720–1725 (2002).
- Sethi, S., Lipford, G., Wagner, H. & Kretzschmar, H. Postexposure prophylaxis against prion disease with a stimulator of innate immunity. *Lancet* **360**, 229–230 (2002).
- Sigurdsson, E. M. *et al.* Immunisation delays the onset of prion disease in mice. *Am. J. Pathol.* **161**, 13–17 (2002).
- Bard, F. *et al.* Peripherally administered antibodies against amyloid β -peptide enter the central nervous system and reduce pathology in a mouse model of Alzheimer disease. *Nature Med.* **6**, 916–919 (2000).
- Brandner, S. *et al.* Normal host prion protein necessary for scrapie-induced neurotoxicity. *Nature* **379**, 339–343 (1996).
- Anstee, D. J. *et al.* New monoclonal antibodies in CD44 and CD58: their use to quantify CD44 and CD58 on normal human erythrocytes and to compare the distribution of CD44 and CD58 in human tissues. *Immunology* **74**, 197–205 (1991).
- Avent, N. D. *et al.* Protein-sequence studies on Rh-related polypeptides suggest the presence of at least two groups of proteins which associate in the human red-cell membrane. *Biochem. J.* **256**, 1043–1046 (1988).
- Dickinson, A. G., Meikle, V. M. & Fraser, H. Identification of a gene which controls the incubation period of some strains of scrapie agent in mice. *J. Comp. Pathol.* **78**, 293–299 (1968).
- Wadsworth, J. D. *et al.* Tissue distribution of protease resistant prion protein in variant Creutzfeldt-Jakob disease using a highly sensitive immunoblotting assay. *Lancet* **358**, 171–180 (2001).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements This work is supported by grants from the Medical Research Council (UK). We thank D. Walsh and S. Gentleman for help with the immunohistochemistry.

Competing interests statement The authors declare competing financial interests: details accompany the paper on *Nature's* website (<http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to S.H. (e-mail: s.hawke@imperial.ac.uk).

Apolipoprotein L-I is the trypanosome lytic factor of human serum

Luc Vanhamme^{*†}, Françoise Paturiaux-Hanocq^{*†}, Philippe Poelvoorde^{*†}, Derek P. Nolan^{*‡}, Laurence Lins[§], Jan Van Den Abbeele^{||}, Annette Pays^{*}, Patricia Tebabi^{*}, Huang Van Xong[¶], Alain Jacquet[#], Nicole Moguilevsky[#], Marc Dieu[☆], John P. Kane^{**}, Patrick De Baetselier[¶], Robert Brasseur[§] & Etienne Pays^{*}

^{*} Laboratory of Molecular Parasitology, IBMM, University of Brussels, 12, rue des Profs Jeener et Brachet, B6041 Gosselies, Belgium

[§] Centre de Biophysique Moléculaire Numérique, University of Gembloux, B5030 Belgium

^{||} Department of Parasitology, Institute for Tropical Medicine, Antwerp, B2000 Belgium

[¶] Laboratory of Cellular Immunology, Free University of Brussels, Sint Genesius Rode, B1640 Belgium

[#] Laboratory of Applied Genetics, IBMM, University of Brussels, B6041 Belgium

[☆] Unit of Cellular Biochemistry & Biology, University of Namur, B5000 Belgium

^{**} Cardiovascular Research Institute, University of California, San Francisco, California, 94143-0130 USA

[†] These authors contributed equally to this work

Human sleeping sickness in east Africa is caused by the parasite *Trypanosoma brucei rhodesiense*. The basis of this pathology is the resistance of these parasites to lysis by normal human serum (NHS)^{1,2}. Resistance to NHS is conferred by a gene that encodes a truncated form of the variant surface glycoprotein termed serum resistance associated protein (SRA)^{3,4}. We show that SRA is a lysosomal protein, and that the amino-terminal α -helix of SRA is

[‡] Present address: Department of Biochemistry, Trinity College, Dublin 2, Ireland.

responsible for resistance to NHS. This domain interacts strongly with a carboxy-terminal α -helix of the human-specific serum protein apolipoprotein L-I (apoL-I). Depleting NHS of apoL-I, by incubation with SRA or anti-apoL-I, led to the complete loss of trypanolytic activity. Addition of native or recombinant apoL-I either to apoL-I-depleted NHS or to fetal calf serum induced lysis of NHS-sensitive, but not NHS-resistant, trypanosomes. Confocal microscopy demonstrated that apoL-I is taken up through the endocytic pathway into the lysosome. We propose that apoL-I is the trypanosome lytic factor of NHS, and that SRA confers resistance to lysis by interaction with apoL-I in the lysosome.

Trypanosoma brucei brucei infects a wide range of mammals but is unable to infect humans because this *T. brucei* subspecies is lysed by NHS^{1,2}. In contrast, *T. b. rhodesiense* exhibits sensitivity or resistance to NHS, depending on antigenic variation. The latter phenotype arises because the gene encoding resistance, SRA, is present in just one of the multiple polycistronic units where the genes for the variant surface glycoprotein (VSG) are transcribed⁴. The product of SRA is an atypical VSG of shorter than average length (410 amino acids instead of approximately 490)³. Anti-SRA antibodies detected a protein with a relative molecular mass of around 50,000 (M_r , 50K; Fig. 1, lanes 1, 2), which accumulates in the lysosome as revealed with a specific marker for the endocytic compartment (tomato lectin, which stains the flagellar pocket, endosomes and lysosome)⁵, and antibodies against the lysosomal membrane glycoprotein p67 (ref. 6; Fig. 2a, b).

Because transfection of SRA allows *T. b. brucei* to grow in NHS⁴ (Fig. 1, lanes 3–5), we investigated the minimal requirements of SRA necessary for resistance by generating trypanosomes expressing various SRA mutants (Fig. 3d). These cells were analysed for SRA expression and sensitivity to NHS (Fig. 1; see also Supplementary Information). Although SRA contains N-linked high mannans (Supplementary Information), these were not involved in resistance, as individual or collective replacement of the putative N-glycosylation sites did not affect SRA-mediated resistance (Fig. 1, lanes 6, 7). The C terminus and glycosyl phosphatidylinositol (GPI) anchoring signal were also dispensable, as truncated SRA versions lacking this region still conferred resistance (Fig. 1, lanes 8, 9). Similarly, deletions of the N-terminal signal peptide or of the most hydrophobic region (amino acids 98–118) were without effect (Fig. 1, lanes 10, 11). Regions between residues 32–97 and/or 119–192 were mapped to be the minimal sections of SRA required for resistance. Because the former region contains distinctive features of the amphipathic α -helix involved in VSG dimerization, for example, heptad repeats⁷, we introduced point mutations predicted to disrupt this helix. Two independent double mutations (L61P/I62P and I58Q/I62Q) abolished the resistance phenotype, whereas mutations in an adjacent region (V40Q/V43Q/L47Q) did not (Fig. 1, lanes 12–14). Thus, the α -helix at residues 54–65 containing the key hydrophobic residues I58, L61 and I62 was necessary to confer resistance.

The finding that an α -helical region similar to the one involved in the dimerization of VSGs was required for SRA-mediated resistance raised the possibility that SRA neutralizes trypanolytic constituents of NHS through a coiled-coil protein–protein interaction. Therefore, we subjected NHS to affinity chromatography using an N-terminal His-tagged version of SRA (residues 1–251) immobilized onto Ni-nitrilotriacetic acid (NTA) agarose. A 40K doublet was specifically retained by SRA (Fig. 3a, lanes 1, 2), and this occurred in various conditions (pH 7.5 or 5.8 and 0 or 0.6 M NaCl). Notably, the doublet was not observed using L61P/I62P SRA—the α -helix-disrupted mutant SRA that does not confer resistance—whereas the doublet was still present with a functional SRA mutant (Fig. 3a, lanes 3, 4). Similar results were obtained using SRA covalently attached to Sepharose. In this case, the doublet was the only component bound, apart from a 65K contaminant (Fig. 3a, lanes 5, 6). Mass spectrometry identified the 40K doublet as apoL-I, a

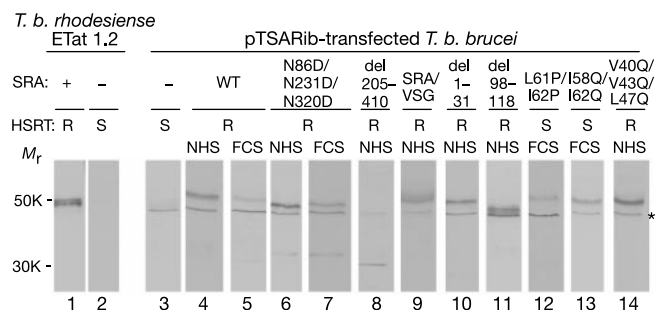


Figure 1 Expression of SRA and phenotype of pTSARib-transfected *T. b. brucei*. The western blots obtained with anti-SRA antibodies are shown for *T. b. rhodesiense* ETat 1.2 clones where SRA is expressed (+) or not (–)⁴, and for various *T. b. brucei* transformants. The pTSARib vector contained or lacked (–), different versions of SRA (WT, wild type; del, deletion of the indicated peptide; SRA/VSG, chimera of the SRA 1–192 and VSG 288–490 peptides). The cells were analysed by the human serum resistance test (HSRT) *in vivo* and *in vitro*⁴, and defined as resistant (R) or sensitive (S). The transformants were grown in mice in the presence of fetal calf serum (FCS) or NHS. The significance of the 45K band (asterisk) is unknown.

human apolipoprotein associated with high density lipoprotein (HDL)^{8–11} (whereas the 65K protein was serum albumin). Anti-apoL-I antibodies confirmed this finding (Fig. 3a, lanes 7, 8).

The apoL-I gene was amplified by polymerase chain reaction with reverse transcriptase (RT–PCR) from RNA of HepG2 cells and cloned into an expression vector. Various mutants of apoL-I were

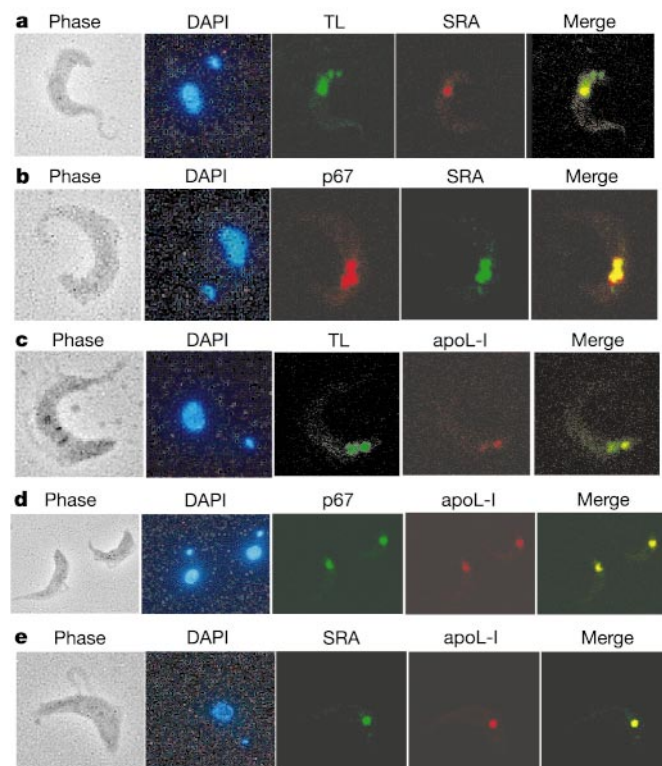


Figure 2 Localization of SRA and apoL-I in the *T. b. rhodesiense* ETat 1.2R clone. The trypanosomes shown in phase contrast were analysed by confocal fluorescence microscopy after incubation for 1 h at 37 °C in the absence (a, b) or presence (c–e) of Alexa 594–apoL-I. DAPI (4,6-diamidino-2-phenylindole), tomato lectin (TL) and anti-p67 monoclonal antibodies (p67) label the DNA (large dot, nucleus; small dot, kinetoplast), endocytic compartment and lysosome, respectively^{5,6}. Merged images show the localization of SRA or apoL-I with respect to the endocytic compartment and lysosome. Serial sections of these pictures are shown in Supplementary Information.

generated, concentrating on the sequence for a peptide (residues 343–355) exhibiting membrane-disrupting activity *in vitro* (Supplementary Information). In particular, we constructed two truncated versions ending at amino acids 342 and 355, respectively, and a mutant with compromised lipid-destabilizing properties (L345Y/L352A/Y354L)¹². [³⁵S]-labelled wild-type and mutant versions of apoL-I were synthesized *in vitro* and incubated with His-tagged SRA. [³⁵S]-labelled apoL-I bound to SRA as did the native protein (Fig. 3a, lane 10). No binding occurred using appropriate control resins (no protein or irrelevant His-tagged protein). Notably, [³⁵S]-labelled apoL-I did not bind to one of the helix-disrupted mutants of SRA (L61P/I62P, lane 9), and showed a significantly reduced binding (50%) to the other mutant (I58Q/I62Q). Deletion of the apoL-I C terminus from residue 343 reduced the binding to 5%, whereas the deletion mutant from residue 356 and the L345Y/L352A/Y354L point mutant showed 24% and 67% binding, respectively. These data indicate that the C-terminal region of apoL-I, which contains an amphipathic α -helix⁸, is important for the interaction with SRA.

The SRA–apoL-I interaction was confirmed by another approach: incubation with V5-tagged apoL-I led to co-immunoprecipitation of SRA by anti-V5 antibodies (Fig. 3b). In addition, a synthetic peptide corresponding to the α -helix of SRA (amino acids 54–65) appeared to interact directly with the 345–355 peptide of apoL-I, as it abolished the liposome fusogenic activity of apoL-I (Supplementary Information). A model of the interacting helices was built using the known structure of the VSG helices as a basis (Fig. 3c, d). The partners would be in anti-parallel configuration

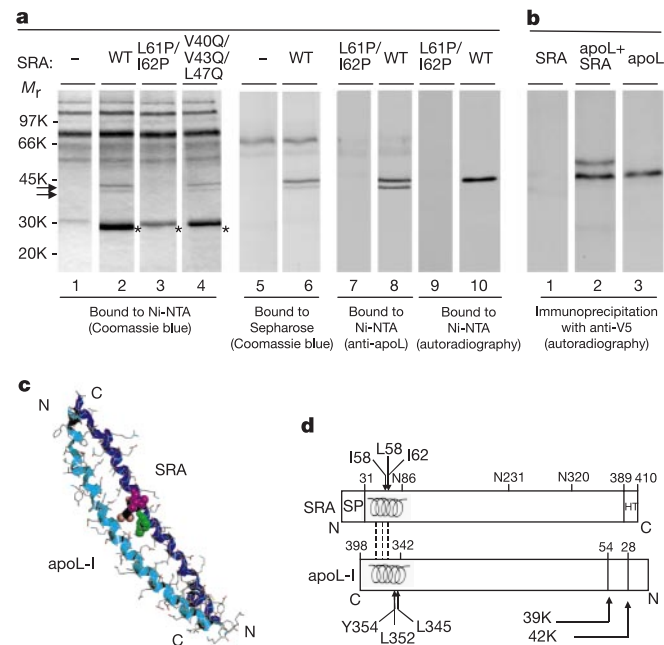


Figure 3 Binding of apoL-I to SRA. **a**, Patterns of NHS proteins bound to resins containing either no protein (–), or different versions of His-tagged SRA. NHS (lanes 1–8) or *in vitro* synthesized [³⁵S]-labelled apoL-I (lanes 9, 10) was incubated with the resin. Bound proteins were eluted with either imidazole (lanes 1–4, 7–10) or deoxycholate (lanes 5, 6), and revealed by Coomassie blue staining (lanes 1–6), binding of anti-apoL-I antibodies (lanes 7, 8) or autoradiography (lanes 9, 10). The double arrow designates bands specifically bound with functional SRA, and asterisks label eluted His-tagged SRA. **b**, Immunoprecipitation with anti-V5 antibodies, of *in vitro* synthesized [³⁵S]-labelled SRA (48K) incubated with or without [³⁵S]-labelled V5–apoL-I (42K). The predicted [³⁵S]-labelled methionine ratio between these proteins is 4:9. **c**, Model of the SRA 31–79 fragment (dark blue ribbon) in anti-parallel interaction with the apoL-I 340–392 fragment (light blue ribbon). I58 (green), L61 (pink) and I62 (mauve) are in CPK representation. **d**, SRA and apoL-I features mentioned in the text (SP, signal peptide; HT, hydrophobic tail; broken lines symbolize the interactions between the α -helices in each molecule).

(–88 and –72 kcal mol^{–1} for the respective hydrophobic energy of the anti-parallel and parallel matchings, to be compared with –89 and –77 kcal mol^{–1} for the interaction between helices A and B in the reference VSG MiTat 1.2 (ref. 7) and in SRA, respectively). The phenotypic effect of the SRA mutants could be explained by differences of complex stability, with –77 and –91 kcal mol^{–1} for the respective hydrophobic energy of the complexes between apoL-I and the I58Q/I62Q and V40Q/V43Q/L47Q SRA mutants, whereas in the L61P/I62P mutant the helical structure is broken and thus no stable complex can be formed.

We investigated the possible involvement of apoL-I in trypanosome lysis. ApoL-I was found to be internalized through the endocytic pathway (Fig. 2c), and to localize with SRA in the lysosome (Fig. 2d, e). Fractionation of NHS on SRA–Sepharose, which essentially retains apoL-I (Fig. 3a, lanes 5, 6), reproducibly

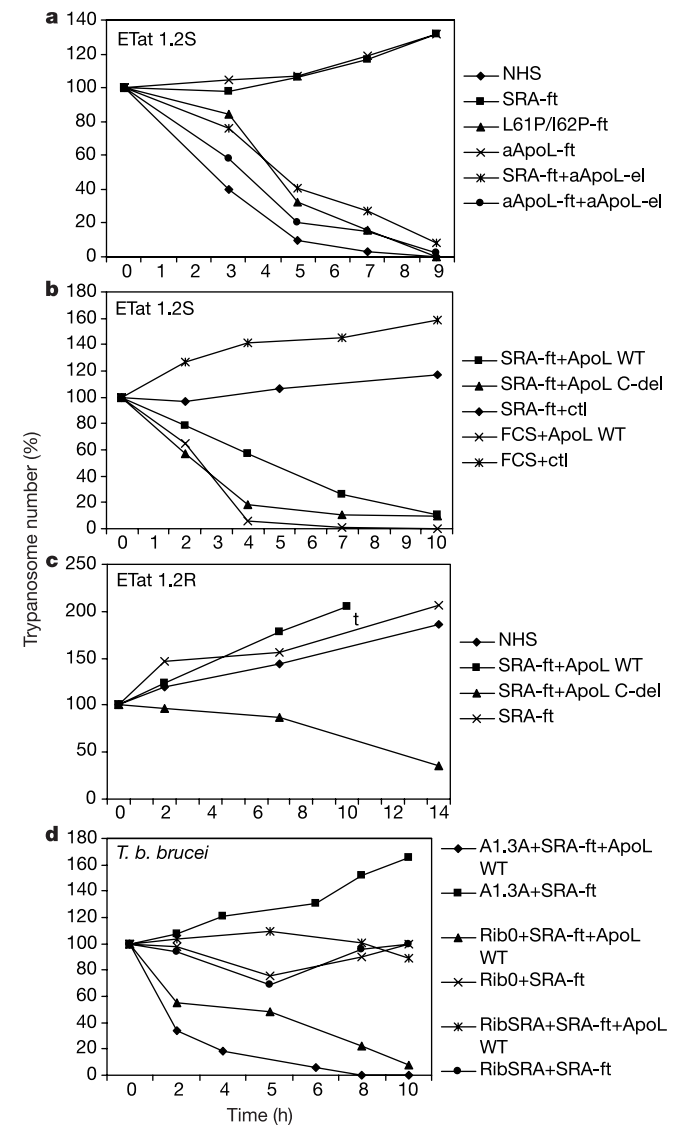


Figure 4 Lytic activity of apoL-I. **a**, Incubation of Etat 1.2S with differently treated NHS (SRA-ft, L61P/I62P-ft, aApoL-ft indicate flow-through fractions from SRA–, L61P/I62P SRA– and anti-apoL-I–Sepharose, respectively; aApoL-el, eluate of the fraction bound to anti-apoL-I–Sepharose). **b**, Incubation of Etat 1.2S in either SRA-ft or FCS supplemented with recombinant apoL-I (C-del, lacking the C-terminal 343–398 peptide) or with the equivalent fraction from control CHO cells (ctl). **c**, Incubation of Etat 1.2R in NHS or SRA-ft supplemented with recombinant apoL-I. **d**, Incubation of different *T. brucei* cell lines in SRA-ft supplemented with or without recombinant apoL-I. The pTSARib transformants (Rib0, RibSRA) are pleomorphic forms that do not grow under these incubation conditions.

resulted in total loss of lytic activity, as tested both *in vitro* (Fig. 4a) and in mice (see Methods). This result was not obtained with resins devoid of SRA, or containing L61P/I62P SRA (Fig. 4a). Lytic activity was also lost when NHS was fractionated on anti-apoL-I coupled to Sepharose (Fig. 4a). As expected this resin retained apoL-I, together with various amounts of serum albumin, apolipoprotein J, amyloid protein P and transthyretin (data not shown). Whereas elution of apoL-I from the SRA column required high detergent concentrations, native apoL-I was easily eluted from anti-apoL-I coupled to Sepharose. Addition of this fraction to sera depleted on either SRA–Sepharose or anti-apoL-I–Sepharose completely restored lytic activity (Fig. 4a). To be certain that apoL-I was responsible for this reconstitution, we expressed recombinant His-tagged apoL-I in Chinese hamster ovary (CHO) cells and added the purified protein to NHS that had been depleted using SRA–Sepharose. The addition of physiological concentrations of recombinant apoL-I (around $8 \mu\text{g ml}^{-1}$)⁸, fully restored the lytic activity on either *T. b. brucei* (AnTat 1.3A and pTSARib-0 transformants) or NHS-sensitive *T. b. rhodesiense* (ETat 1.2S clone), whereas an equivalent extract from control CHO cells did not (Fig. 4b, d). Notably, the same lytic effect of apoL-I was obtained if fetal calf serum was used instead of apoL-I-depleted NHS (Fig. 4b), and lysis was not observed on NHS-resistant cells (either *T. b. rhodesiense* ETat1.2R or *T. b. brucei* pTSARib-SRA) (Fig. 4c, d). Under the same conditions, recombinant apoL-I lacking the 343–398 C-terminal peptide still induced full lysis of NHS-sensitive trypanosomes, but also affected NHS-resistant cells (Fig. 4b, c). Such an effect by an apoL-I mutant deficient in interaction with SRA confirmed that this interaction is involved in resistance to lysis.

Our data show that apoL-I is the trypanolytic factor of NHS and that SRA neutralizes its activity, presumably through a coiled-coil protein–protein interaction. Previous studies identified the lytic factor of NHS as haptoglobin-related protein (Hpr), another HDL-linked serum protein found only in primates¹³, but this view was debated^{2,14}. Moreover, the suggestion that Hpr might be necessary to allow uptake of the lytic particles into trypanosomes¹⁵, is contradicted by our observation that apoL-I kills trypanosomes in Hpr-free medium. The mechanism of lysis by apoL-I is not understood. Indeed, despite the evidence that lysis by NHS is due to disruption of the lysosomal membrane^{1,16}, our data show that the fusogenic peptide (residues 343–355) of apoL-I is not required. Resistance to lysis seems to be due to SRA-mediated inhibition of the apoL-I lytic effect within the lysosome. This work opens the way for the characterization of peptides that would be useful for both therapy of sleeping sickness due to *T. b. rhodesiense* and prevention of nagana in cattle. □

Methods

Generation of transgenic trypanosomes expressing SRA mutants

The SRA gene in pTSARib-SRA⁴ was altered by site-directed mutagenesis (Stratagene) and/or restriction endonuclease digestion and ligation with fragments from the AnTat 11.17 VSG¹⁷. The resultant plasmids were electroporated into *T. b. brucei* procyclic forms that were cyclically transmitted through tsetse flies to obtain bloodstream form transformants⁴.

Production of SRA

SRA was produced either by *in vitro* RNA translation using BSDP1400HA65-SRA¹⁸, or by expression in *Escherichia coli* using pQE30 in which the 1–753-base-pair SRA fragment was inserted between BamHI and HindIII sites to encode an N-terminal, His-tagged polypeptide. In the latter case, the protein was solubilized from pellets of sonicated cell extracts, by incubation for 1 h at room temperature in 8 M urea, 2 mM mercaptoethanol, 0.1 M NaH₂PO₄, 10 mM Tris (pH 8), then dialysis against the same buffer containing successively 0.5 M urea, 0.35 M urea, 0.1 M urea and 1% CHAPS buffer.

Antibodies

The anti-SRA antibodies were generated by rabbit immunization with the His-tagged 1–251 polypeptide, and affinity purified by binding to antigen-coated nitrocellulose filters and acid elution. They were used after 1:2 and 1:100 dilution, for immunofluorescence and western blot analysis, respectively. The anti-apoL-I antibodies⁸ were used at 1:20,000 dilution. Anti-p67 antibodies were from the Ali I-218 hybridoma supernatant (gift of D. G. Russell).

Interaction of SRA with NHS

Recombinant wild type or mutant His-tagged SRA (100 μg) was incubated with 500 μl NHS for 4 h at 4 °C in 0.6 M NaCl, 0.35% CHAPS buffer, 0.15 M MES buffer (pH 5.8) with a cocktail of protease inhibitors (complete EDTA-free including pepstatin, Roche) (buffer A). The mixture was then incubated for 30 min at 4 °C with 100 μl Ni–NTA beads (Qiagen). Processing of these beads, including elution of bound material with 250 mM imidazole, was performed in buffer A as described by Qiagen. Alternatively, His-tagged SRA was covalently coupled to activated CH Sepharose 4B (Pharmacia), and the bound material was eluted in 5% Na deoxycholate, 50 mM Tris (pH 7.5).

Confocal fluorescence microscopy

The trypanosomes were fixed for 10 min in 3.7% paraformaldehyde and permeabilized for 10 min in 0.1% Triton X-100. Biotinylated tomato lectin, Alexa 488-streptavidin and Alexa 488/594-anti-mouse or anti-rabbit secondary antibodies were used. To detect apoL-I, the deoxycholate eluate from SRA–Sepharose was coupled to Alexa 594 after extensive dialysis and depletion of serum albumin by gel filtration, and incubated with trypanosomes for 1 h at 37 °C. The samples were analysed with a Leica TCS SP2 confocal microscope.

Production of native apoL-I

ApoL-I was obtained by elution of the NHS material bound to anti-apoL-I–Sepharose, using 5% CHAPS, 0.1 M glycine (pH 2.8) followed by neutralization with 1 M Tris (pH 8.0). Residual contaminants were identified by mass spectrometry (see text).

Cloning of apoL-I

The apoL-I gene was amplified by RT-PCR using total RNA from HepG2 cells and the following pair of oligonucleotides: ApoL-F: 5'-TGTCCTCTGCGGTACCATG AGTGCACCTTTTCCTTGGTGTGAGAG-3'; ApoL-R: 5'-CCCTGCCCTGCTCGAGCAGT TCTTGTCGCCCTGCGAATC-3'. The 1.15-kilobase fragment was digested by *KpnI* and *XhoI*, and ligated with *KpnI/XhoI*-digested pcDNA3.1/V5-HisA plasmid (Invitrogen). Mutants were generated by site-directed mutagenesis (Stratagene) and by deletions and ligations to remove specific fragments.

Interaction of SRA with *in vitro* synthesized apoL-I

[³⁵S]-labelled apoL-I was produced by *in vitro* translation in reticulocyte lysates, using the apoL-I-pcDNA3.1/V5-His construct or mutant derivatives as template for transcription. All apoL-I polypeptides contained an in-frame V5 tag at their C terminus, and some were additionally flanked with a C-terminal His tag, depending on the insertion of an in-frame stop codon between the two tag sequences. For interaction with His-tagged SRA, apoL-I with the V5 tag alone was used. In this case, 5 μl reticulocyte lysate containing [³⁵S]-labelled apoL-I was incubated for 4 h at 4 °C with 15 μg His-tagged SRA in buffer A. Further treatment with Ni–NTA beads was performed as described above. The bound/unbound apoL-I was revealed by autoradiography and quantified by liquid scintillation counting after precipitation in 10% trichloroacetic acid.

Modelling

A three-dimensional model of the complex between the apoL-I C-terminal and SRA N-terminal peptides was built using the Swissmodel and Swiss-PDB Viewer programs¹⁹. The three-dimensional structure of the coiled-coil domain (residues 7–112) of the MiTat 1.2 VSG⁷ (Protein data Bank code 2VSG) was used as template. We assumed that the SRA N-terminal and the apoL-I C-terminal fragments interact like the two VSG helices. Residues 31–79 of SRA were aligned to residues 7–54 of the mature VSG (first helix), as that SRA fragment is the most similar to the latter domain of the VSG. ApoL-I residues 340–392 were therefore aligned to residues 59–112 of the VSG (second helix). The parallel (N–C versus N–C) and anti-parallel (N–C versus C–N) matchings were built. The resulting structures were minimized using Hyperchem 5.0 (Hypercube, Inc.), with the conjugate gradient method and AMBER force field. The lowest energy structure was kept. The stereochemical quality of the model was checked with Procheck²⁰: 98% of the residues were in the allowed regions of the Ramachandran plot. Molecular views were drawn with the WinMGM program²¹.

Dissociation/reconstitution of NHS trypanolytic activity

Untreated, affinity-depleted and/or supplemented serum was added to HMI-9 medium containing 5% CHAPS, to a final concentration of 25%. Recombinant apoL-I was used at physiological concentration (8 $\mu\text{g ml}^{-1}$). After a 2 h incubation, CHAPS was removed by extensive dialysis. The trypanosomes were seeded at 5×10^5 cells ml^{-1} . The results shown are representative of at least three separate experiments, and they were confirmed *in vivo*: when directly injected into NMRI mice, the incubation mixtures tested as non-lytic or lytic led to detectable parasitaemia after 2 days or at least 10 days, respectively.

Received 15 November 2002; accepted 3 February 2003; doi:10.1038/nature01461.

- Hager, K. M. *et al.* Endocytosis of a cytotoxic human high density lipoprotein results in disruption of acidic intracellular vesicles and subsequent killing of African trypanosomes. *J. Cell Biol.* **126**, 155–167 (1994).
- Raper, J., Portela, M. P., Lugli, E., Frevert, U. & Tomlinson, S. Trypanosome lytic factors: novel mediators of human innate immunity. *Curr. Opin. Microbiol.* **4**, 402–408 (2001).
- De Greef, C. & Hamers, R. The serum-resistance associated (SRA) gene of *Trypanosoma brucei rhodesiense* encodes a VSG-like protein. *Mol. Biochem. Parasitol.* **68**, 277–284 (1994).
- Xong, H. V. *et al.* A VSG expression site-associated gene confers resistance to human serum in *Trypanosoma rhodesiense*. *Cell* **95**, 839–846 (1998).
- Nolan, D. P., Geuskens, M. & Pays, E. N-linked glycans containing linear poly-N-acetylactosamine as sorting signals in endocytosis in *Trypanosoma brucei*. *Curr. Biol.* **9**, 1169–1172 (1999).
- Kelley, R. J., Alexander, D. L., Cowan, C., Balber, A. E. & Bangs, J. D. Molecular cloning of p67, a

- lysosomal membrane glycoprotein from *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **98**, 17–28 (1999).
7. Blum, M. L. *et al.* A structural motif in the variant surface glycoproteins of *Trypanosoma brucei*. *Nature* **362**, 603–609 (1993).
 8. Duchateau, P. N. *et al.* Apolipoprotein L, a new human high density lipoprotein apolipoprotein expressed by the pancreas. Identification, cloning, characterization, and plasma distribution of apolipoprotein L. *J. Biol. Chem.* **272**, 25576–25582 (1997).
 9. Page, N. M., Butlin, D. J., Lomthaisong, K. & Lowry, P. J. The human apolipoprotein L gene cluster: identification, classification, and sites of distribution. *Genomics* **74**, 71–78 (2001).
 10. Duchateau, P. N., Pullinger, C. R., Cho, M. H., Eng, C. & Kane, J. P. Apolipoprotein L gene family: tissue-specific expression, splicing, promoter regions; discovery of a new gene. *J. Lipid Res.* **42**, 620–630 (2001).
 11. Monajemi, H., Fontijn, R. D., Pannekoek, H. & Horrevorts, A. J. The apolipoprotein L gene cluster has emerged recently in evolution and is expressed in human vascular tissue. *Genomics* **79**, 539–546 (2002).
 12. Lins, L., Charlotiaux, B., Thomas, A. & Brasseur, R. Computational study of lipid-destabilizing protein fragments: towards a comprehensive view of tilted peptides. *Proteins* **44**, 435–447 (2001).
 13. Smith, A. B., Esko, J. D. & Hajduk, S. L. Killing of trypanosomes by the human haptoglobin-related protein. *Science* **268**, 284–286 (1995).
 14. Hataida, S. *et al.* No trypanosome lytic activity in the sera of mice producing human haptoglobin-related protein. *Mol. Biochem. Parasitol.* **119**, 291–294 (2002).
 15. Drain, J., Bishop, J. R. & Hajduk, S. L. Haptoglobin-related protein mediates trypanosome lytic factor binding to trypanosomes. *J. Biol. Chem.* **276**, 30254–30260 (2001).
 16. Shimamura, M., Hager, K. M. & Hajduk, S. L. The lysosomal targeting and intracellular metabolism of trypanosome lytic factor by *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **115**, 227–237 (2001).
 17. Do Thi, C. D., Aerts, D., Steinert, M. & Pays, E. High homology between variant surface glycoprotein gene expression sites of *Trypanosoma brucei* and *Trypanosoma gambiense*. *Mol. Biochem. Parasitol.* **48**, 199–210 (1991).
 18. Salmon, D. *et al.* A novel heterodimeric transferrin receptor encoded by a pair of VSG expression site-associated genes in *T. brucei*. *Cell* **78**, 75–86 (1994).
 19. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 14–23 (1997).
 20. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
 21. Brasseur, R. WinMGM: a fast CPK molecular graphics program for analyzing molecular structure. *J. Mol. Graphics* **12**, 212–218 (1994).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank M. Chamekh for the pQ30-SRA plasmid, A. Michel and V. Stroobant for peptide synthesis, S. Kler and C. Flore for technical assistance, and D. Perez-Morza and S. Schurmans for help. This work was supported by the Belgian Fund for Scientific Research (FRSM, FWO and Crédit aux Chercheurs) and by the Interuniversity Attraction Poles Programme, Belgian State, Federal Office for Scientific, Technical and Cultural Affairs. R.B., L.V. and L.L. are, respectively, Research Director and Research Associates at the Belgian Fund for Scientific Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.P.
(e-mail: epays@ulb.ac.be).

Regulated degradation of a class V myosin receptor directs movement of the yeast vacuole

Fusheng Tang, Emily J. Kauffman, Jennifer L. Novak, Johnathan J. Nau, Natalie L. Catlett* & Lois S. Weisman

Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242, USA

Normal cellular function requires that organelles be positioned in specific locations. The direction in which molecular motors move organelles is based in part on the polarity of microtubules and actin filaments^{1–3}. However, this alone does not determine the intracellular destination of organelles. For example, the yeast class V myosin, Myo2p, moves several organelles to distinct locations during the cell cycle^{4–8}. Thus the movement of each

type of Myo2p cargo must be regulated uniquely. Here we report a regulatory mechanism that specifically provides directionality to vacuole movement. The vacuole-specific Myo2p receptor, Vac17p, has a key function in this process. Vac17p binds simultaneously to Myo2p and to Vac8p, a vacuolar membrane protein. The transport complex, Myo2p–Vac17p–Vac8p, moves the vacuole to the bud, and is then disrupted through the degradation of Vac17p. The vacuole is ultimately deposited near the centre of the bud. Removal of a PEST sequence (a potential signal for rapid protein degradation) within Vac17p causes its stabilization and the subsequent 'backward' movement of vacuoles, which mistargets them to the neck between the mother cell and the bud. Thus the regulated disruption of this transport complex places the vacuole in its proper location. This may be a general mechanism whereby organelles are deposited at their terminal destination.

Genetic studies indicate that Vac8p, a vacuole membrane protein, is required for Myo2p function in vacuole inheritance⁹. However, there was no evidence for a direct interaction between Myo2p and Vac8p. We find here that the protein Vac17p (*Saccharomyces cerevisiae* open reading frame YCL063w), links Myo2p to Vac8p. Vac17p was obtained by several approaches including: (1) identification of the wild-type gene of a newly identified vacuole inheritance mutant *vac17-1* (Fig. 1a); (2) identification of a vacuole-specific, Myo2p-interacting protein¹⁰; and (3) identification of Vac8p-interacting proteins¹¹ (and data not shown). We found that the *vac17Δ* mutant was defective in vacuole inheritance (Fig. 1b, c). Thus, Vac17p has an important role in this process.

Several lines of evidence demonstrate that Vac17p functions as the vacuole-specific receptor for Myo2p. First, using immunofluorescence microscopy, Vac17p was found on the vacuole membrane (Fig. 1e). Consistent with this localization, fractionation studies (Fig. 1h) showed that Vac17p was enriched on isolated vacuoles. Second, Vac17p seems to be required only for vacuole inheritance; deletion of *VAC17* did not affect the movement of other Myo2p cargoes¹⁰ and did not show any growth defects (data not shown). Third, Vac17p interacts with Myo2p. Anti-Vac17p antibody precipitated Myo2p (Fig. 1i) and anti-Myo2p antibody precipitated Vac17p (Fig. 1j) from a solubilized crude cell extract. Moreover, Vac17p interacts with Myo2p in a yeast two-hybrid assay¹⁰. Fourth, the Myo2p–Vac17p interaction is essential for vacuole inheritance. Deletion of the Myo2p-binding region (residues 109–190) of Vac17p blocks vacuole inheritance (Fig. 2c). Furthermore, *myo2* mutants that are specifically defective in vacuole inheritance (*myo2-2*, *myo2-N1304S*) do not interact with Vac17p¹⁰. Both the defect in vacuole inheritance and the failure of these two *myo2* mutants to interact with Vac17p are suppressed by a point mutation in the Myo2p-binding domain of Vac17p (Vac17p-I140V)¹⁰. These latter findings indicate that Myo2p and Vac17p directly interact with each other.

We postulated that Vac17p, as the vacuole-specific Myo2p receptor, must attach to the vacuole membrane to function. Vac17p does not appear to contain any transmembrane domains (Fig. 2a). We found that Vac17p–Vac8p interactions are required for normal Vac17p association with the vacuole. In the *vac8Δ* mutant, a significantly smaller portion of Vac17p is associated with the vacuole membrane and much of the Vac17p had an altered location (green dots in Fig. 1f). Moreover, Vac17(Δ319–377)p, which is missing the Vac8p-binding domain, has the same defect in its localization (green dots in Fig. 1g). Similar results were obtained with an *in vitro* analysis of Vac17p distribution. Equal amounts of protein from total cell extracts and isolated vacuoles were separated by SDS–polyacrylamide gel electrophoresis (PAGE) and probed with anti-Vac17p antibody. In the wild-type cells, Vac17p was enriched tenfold in the vacuole fraction. In contrast, in *vac8Δ*, this enrichment was only twofold (Fig. 1h). We noted that both *vac8Δ* and *vac17(Δ319–377)* mutants have greatly elevated levels of

* Present address: Torrey Mesa Research Institute, Syngenta Research and Technology, 3115 Merryfield Row 100, San Diego, California 92121, USA

Vac17p, raising the possibility that the displacement of Vac17p might be independent of a defect in Vac17p–Vac8p interactions, and instead could be due to saturation of an as yet unidentified Vac17p receptor. However, several mutants including *myo2-2*, *myo2-N1304S* and *vac17(Δ97–259)* have similar elevated levels of Vac17p (Fig. 2c, d), yet these mutants retain Vac17p on the vacuole membrane (Fig. 4b). In these mutants both Vac8p and the Vac8p-binding domain of Vac17p are present. Together these findings indicate that Vac8p has an important role in the association of Vac17p with the vacuole membrane.

Co-immunoprecipitation assays demonstrated that Vac17p was in a complex with Vac8p (Fig. 1i). Moreover, Vac17p interacted with Vac8p in a yeast two-hybrid assay (Fig. 2b). In further support of the importance of Vac17p–Vac8p interactions, deletion of the Vac8p-binding domain (pVAC17(Δ319–377)) from Vac17p blocked vacuole inheritance (Fig. 2c). Therefore, Vac17p–Vac8p interactions are required for vacuole inheritance.

Using yeast two-hybrid analysis, we found that the Vac8p-binding domain of Vac17p encompasses residues 290–380 (Fig. 2a, b). Notably, this region of Vac17p contains two LSSL motifs. These motifs were first identified in E-cadherin and have been shown by crystallographic studies to bind directly to specific armadillo repeats in β-catenin¹². Notably, we found in a yeast two-hybrid test that the LSSL region of Vac17p interacts with an armadillo repeat on Vac8p (not shown). Accordingly, we predict that the interaction between Vac17p and Vac8p is direct.

We noted that the Vac8p-binding domain of Vac17p (residues 290–380) is distinct from the Myo2p-binding domain: residues 109–170 (ref. 10) (Fig. 2a). Notably, Vac17p binds Vac8p and Myo2p simultaneously. Immunoprecipitation of Myo2p co-precipitated Vac8p and Vac17p (Fig. 1j); neither protein was co-

precipitated in the *vac17Δ* mutant (Fig. 1j). Therefore, Vac17p tethers Myo2p to Vac8p and thus to the vacuole membrane.

As predicted, mutations that disrupt either Myo2p–Vac17p or Vac17p–Vac8p interactions blocked vacuole inheritance. Unexpectedly, the block in vacuole inheritance resulted in the accumulation of Vac17p (Fig. 2c, d). This observation suggested that the steady-state levels of Vac17p are regulated. Furthermore, this finding raised the possibility that variation of Vac17p levels may control the assembly and disassembly of the Myo2p–Vac17p–Vac8p complex and thereby control vacuole inheritance. The regulation of Vac17p levels could be mediated through control of its expression, its turnover, or both.

Two studies showed that the *VAC17* messenger RNA levels oscillate in coordination with the cell cycle^{13,14}. Consistent with these observations, we found, by means of two independent assays, that Vac17p levels also increased and decreased with the cell cycle. Cells were synchronized by α-factor arrest and release (Fig. 3a). Under these conditions, the first cell cycle (from bud emergence to bud emergence) required 120 min. Vac17p began to increase at 50 min, reached a peak at 70 min, and then decreased at 100 min (Fig. 3a). Vac17p levels were greatest at the same time or slightly later than the time of maximal *VAC17* mRNA levels, which occurred 50 min after α-factor arrest¹³.

The levels of Vac17p on the vacuole membrane also undergo a similar cell-cycle-dependent change. We assessed Vac17p levels in asynchronous populations by monitoring Vac17p on the vacuole membrane using immunofluorescence microscopy. The ratio of the diameter of the bud to the mother cell (the bud–mother index, BMI) serves to indicate the approximate position in the cell cycle. The levels of Vac17p on the vacuole membrane were reported as the intensity of green fluorescence (anti-Vac17p) relative to the red

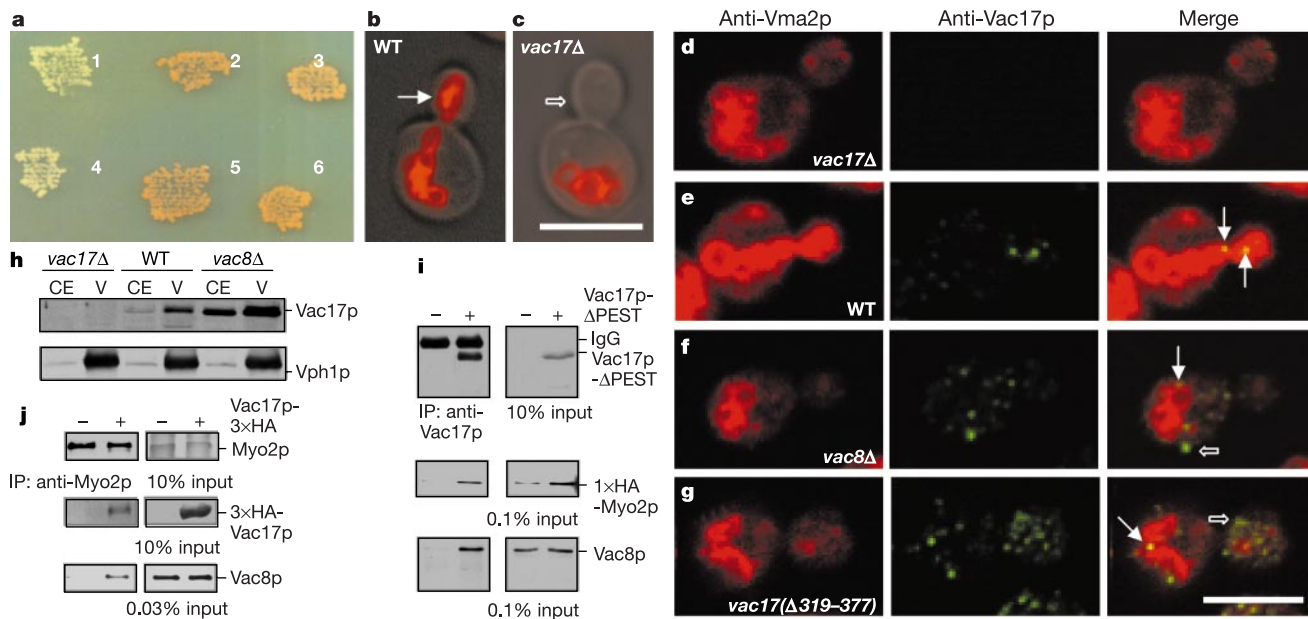


Figure 1 Vac17p is required for vacuole inheritance and forms a complex with Myo2p and Vac8p. **a–c**, YCL063w is the *VAC17* gene. **a**, Patches of cells indicating normal vacuole inheritance (orange) or a defect in vacuole inheritance (yellow) are shown. *vac17-1* (top row) or *vac17Δ* (bottom row) with pRS415 (vector; 1 and 4); the original library plasmid that complements *vac17*, which contains 5 including YCL063w (2 and 5); or pYCL063w-*LEU2* (3 and 6). **b, c**, *vac17Δ* has a vacuole inheritance defect. The filled arrow indicates a bud with inherited vacuoles (**b**), whereas the open arrow indicates a bud lacking inherited vacuoles (**c**). **d–g**, Double-label immunofluorescence of the vacuole membrane (rhodamine red, left), Vac17p (Alexa 488, green, middle) and the merged image (right). **d**, *vac17Δ* (LWY 5798); **e**, wild type (LWY 7235); **f**, *vac8Δ* (LWY 2887); **g**, *vac17(Δ319–377)* (pVAC17(Δ319–377) in LWY 5798). Vac17p is seen as small yellow dots on the vacuole membrane (arrows). Open arrows indicate examples of Vac17p

not on the vacuole membrane (green). Scale bar, 5 μm. **h**, Vac17p co-fractionates with the vacuole. Vacuoles were analysed for Vac17p and Vph1p by western blot analysis. CE, crude cell extracts (20 μg); V, isolated vacuoles (20 μg). The relative intensities of Vac17p estimated with MetaMorph, from left to right, are: 0, 0, 1, 10, 27, 47. **i**, Myo2p and Vac8p immunoprecipitate together with sheep anti-Vac17p antibody. Crude cell extracts from *vac17Δ* (p1 × HA-MYO2, –) or *vac17Δ* (p1 × HA-MYO2, pVAC17-ΔPEST, +) were precipitated with anti-Vac17p antibody and probed for 1 × HA-MYO2p and Vac8p. **j**, The ability of anti-Myo2p antibody to co-precipitate Vac8p requires the presence of Vac17p. Crude cell extracts from *vac17Δ* (–) or *vac17Δ* (p3 × HA-VAC17, +) were precipitated with anti-Myo2p antibody and probed for 3 × HA-Vac17p and Vac8p. IP, immunoprecipitation⁹.

fluorescence (anti-Vma2p). As shown in Fig. 3b, in unbudded wild-type cells (G1 phase) the ratio of Vac17p to vacuole membrane was 0.02. This level increased to 0.07 for cells with a nascent bud (BMI 0.1–0.3) when vacuole inheritance begins. As the bud size increased, the levels of Vac17p decreased (from 0.07 in BMI 0.1–0.3 to 0.04 in BMI 0.5–0.8). When cells entered cytokinesis, and vacuole inheritance was complete (BMI 0.8–1.0), Vac17p decreased further to 0.025.

Analysis of the Vac17p sequence with the PESTfind program predicted a PEST sequence: a potential signal for rapid protein degradation¹⁵ (residues 204–250 of Vac17p; Fig. 2a). Indeed, deletion of this PEST sequence produced elevated levels of Vac17p (pVAC17(Δ 196–251); Fig. 2c). Analysis of the protein levels of VAC17- Δ PEST at specific points in the cell cycle demonstrated

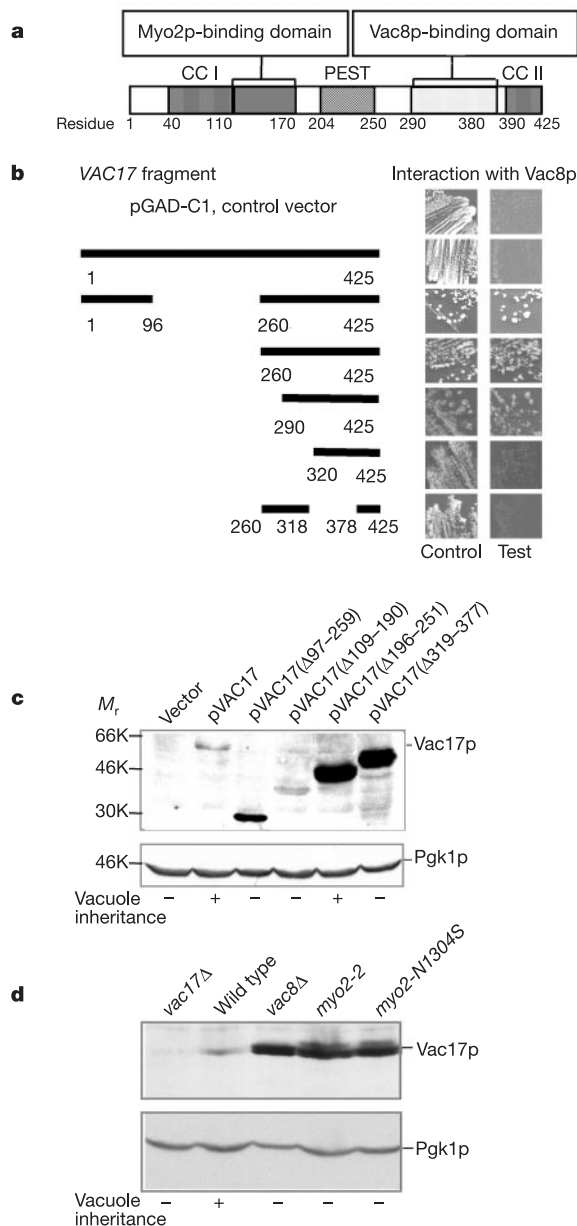


Figure 2 A block in vacuole inheritance or deletion of the Vac17p PEST sequence results in elevated levels of Vac17p. **a**, Diagram of Vac17p. CC, coiled-coil domains. **b**, The indicated regions of Vac17p were fused with the activation domain of Gal4p. Vac8p was fused with the binding domain of Gal4p. Growth on test plates indicates protein–protein interactions. **c**, Western blot analysis of Vac17p and Pgk1p in wild-type and *vac17* mutants. For vacuole inheritance, + indicates more than 85% of the buds contained inherited vacuoles; – indicates less than 5%. **d**, Vac17p levels in selected vacuole inheritance mutants.

that this putative PEST sequence is a rapid degradation signal (Fig. 3b). The large decrease of Vac17p from medium-budded cells (BMI 0.3–0.5) to cells undergoing cytokinesis (BMI 0.8–1.0) in wild-type cells was blocked in VAC17- Δ PEST mutants. Vac17p- Δ PEST levels remained elevated after vacuole inheritance (0.2 with a BMI 0.5–0.8, and 0.17 with a BMI 0.8–1.0). Of note, Vac17p- Δ PEST supported vacuole inheritance to the same extent as wild-type Vac17p (Fig. 2c).

The elevation of Vac17p levels in vacuole inheritance mutants with full-length Vac17p could potentially be caused by an upregulation of Vac17p synthesis, a defect in Vac17p degradation, or both. If blocking vacuole inheritance increases the expression of VAC17, such an upregulation would probably be observed at the time that wild-type cells undergo vacuole inheritance; that is, cells with a BMI between 0.1–0.5 (ref. 16). However, at these bud/mother diameter ratios, the levels of Vac17p in the vacuole inheritance mutant, *myo2-2*, were similar to wild-type cells. In fact, the elevated steady-state levels of Vac17p in *myo2-2* mutants were most divergent from wild-type levels in cells with a BMI of 0.8–1.0, the stage where vacuole inheritance in wild-type cells is complete and Vac17p has significantly decreased, owing to its degradation (Fig. 3b). Thus, it is most likely that the elevated level of Vac17p in *myo2-2* is due to a defect in the degradation of Vac17p.

Although both VAC17- Δ PEST and *myo2-2* mutants have elevated levels of Vac17p, the protein accumulates in different sites. Vac17p missing the PEST sequence accumulated specifically in the bud (Fig. 4a), consistent with Vac17p turnover occurring in this location. In contrast, vacuole inheritance mutants accumulated Vac17p in the mother cell (Fig. 4b). A quantitative analysis of the immunofluorescence data is shown in Fig. 4c. The differential localization of Vac17p in VAC17- Δ PEST versus *myo2-2* is consistent with the hypothesis that the PEST-related degradation of Vac17p occurs specifically in the bud, and that the accumulation of Vac17p in the vacuole inheritance mutants (Fig. 2c, d) is due to the inability of the vacuole-associated Vac17p to enter the bud, where its turnover normally occurs. Together, these observations suggest that the rapid turnover of Vac17p occurs after the vacuole arrives in the bud.

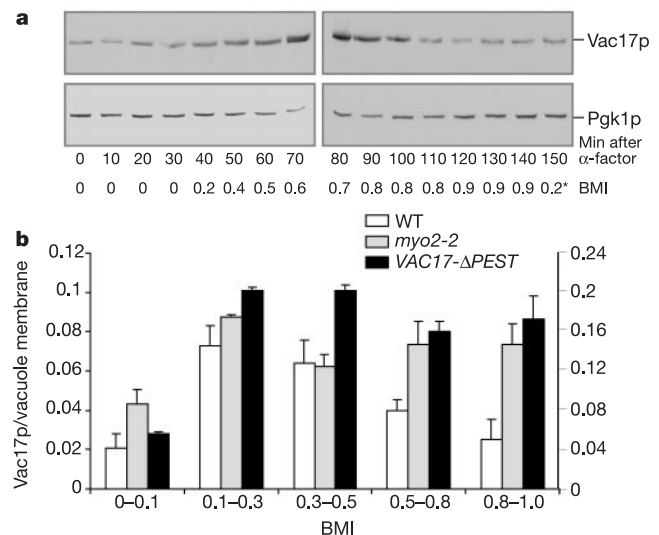


Figure 3 Vac17p levels oscillate during the cell cycle. **a**, Western blot analysis of Vac17p levels in synchronized cells. The BMI was averaged from 50 to 60 cells. An asterisk indicates the BMI of cells with a second round of budding. **b**, Vac17p detected by immunofluorescence microscopy was reported as the ratio of green over red fluorescence in areas (7 μ m²) with yellow dots (measured with Metamorph). The same area was also measured in cells without visible Vac17p. Three independent experiments were averaged and plotted (50–100 cells per category). The left y-axis indicates wild type and *myo2-2* values, the right y-axis indicates Vac17p- Δ PEST levels.

In small-budded cells, Vac17p- Δ PEST generally accumulates at the bud tip (filled arrowhead in Fig. 4a); however, in selected large-budded cells, Vac17p- Δ PEST accumulates at the mother–bud neck (open arrow in Fig. 4a). This localization is reminiscent of the localization of Myo2p. In wild-type cells, Myo2p accumulates at sites of polarized growth, namely, the presumptive bud site, the bud tip in small-budded cells and the mother–bud neck at cytokinesis¹⁷ (reviewed in ref. 18). The localization of Myo2p to the mother–bud neck is thought to occur because Myo2p delivers secretory vesicles to the site of cytokinesis. Notably, while vacuoles move to the presumptive bud site and tips of small-budded cells, vacuoles do not move to the mother–bud neck at cytokinesis (for example, see top panels in Fig. 4d). In a wild-type strain, the vacuoles move away from the mother–bud neck and then do not return to this location.

The stabilization of Vac17p- Δ PEST results in the accumulation of Myo2p on the vacuole membrane¹⁰. This suggests that there is a persistent attachment of Myo2p to the VAC17- Δ PEST vacuole and

raises the possibility that the vacuole may be ‘dragged’ to all intracellular sites that are destinations for Myo2p. Indeed, in a population ($N = 99$) of VAC17- Δ PEST cells with a BMI of 0.8–1, we found that 30% had vacuoles at the mother–bud neck (see filled arrowhead in Fig. 4d, bottom panels), a location not observed in wild-type cells ($N = 75$). We then collected time courses of individual cells to determine whether the vacuoles at the mother–bud neck were trapped on their way into the bud, or whether they had indeed moved backward from the bud centre to the mother–bud neck. We analysed cells over a period of 40 min. To score cells at the end of the cell cycle, the first appearance of a new bud on the mother cell was defined as 0 time, with the previous time points reported relative to the appearance of the new bud (Fig. 4d). In 13 out of 13 wild-type cells, the vacuoles moved away from the mother–bud neck and in the last 20 min showed little change in their position. In contrast, in 9 out of 10 VAC17- Δ PEST cells, the vacuoles moved back to the mother–bud neck. Thus, the regulated degradation of

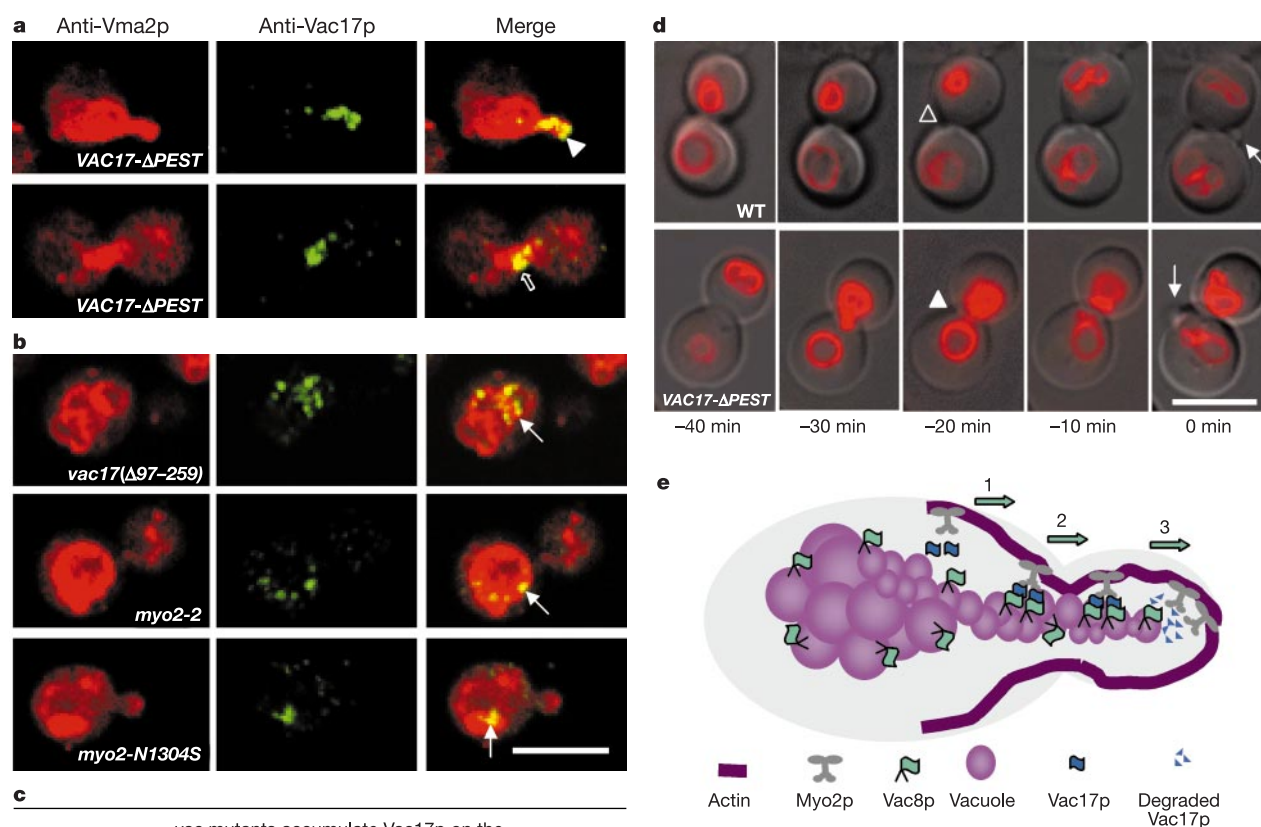


Figure 4 Regulated degradation of Vac17p releases Myo2p and deposits the vacuoles in the bud. **a**, Vac17p- Δ PEST levels are significantly enriched on the vacuole membrane at regions closest to the bud tip (arrowhead) and in the mother–bud neck (open arrow). **b**, Vacuole inheritance mutants accumulate Vac17p on the vacuole membrane in the mother cell (arrows). Scale bar, 5 μ m. **c**, The presence of Vac17p in the mother, the bud, or both, in wild-type and the indicated mutant cells. **d**, VAC17- Δ PEST causes the

movement of vacuoles backward to the mother–bud neck. Pictures were taken every 10 min for a period of 40 min. The nascent buds (arrows), a mother–bud neck devoid of vacuoles (open arrowhead), and vacuoles that have moved back to the mother–bud neck (filled arrowhead) are indicated. **e**, Model for Vac17p function in vacuole inheritance. Because each Myo2p dimer contains two globular tail domains, Vac17p may attach to both domains; however, the stoichiometry of this complex is unknown.

Vac17p in the bud at a specific time in the cell cycle contributes to the directional movement of the vacuole. Moreover, it appears that the regulated degradation of Vac17p releases Myo2p from the vacuole and frees Myo2p to attach to secretory vesicles that are then moved to the site of cytokinesis.

The fact that Vac17p is specifically degraded in the bud was not anticipated. So far, no known bud-specific degradation machinery has been discovered. Although it is possible that such a degradation complex exists, it is not an absolute requirement for achieving bud-specific degradation of Vac17p. The degradation machinery may be present throughout the cytoplasm, or within the vacuole, but the signal for Vac17p turnover could be manifested specifically in the bud through modification of Vac17p. For instance, proteins such as Cbk1p (a kinase) and its activator Mob2p localize specifically to the bud^{19,20}.

Our results support the following model (Fig. 4e). The newly synthesized Vac17p in the mother cell attaches to the vacuole membrane by means of Vac8p and assembles the transport complex, Myo2p–Vac17p–Vac8p (step 1). This transport complex moves the attached vacuole membrane along actin cables into the bud (step 2). After the arrival of vacuole membranes in the bud, Vac17p is degraded, thereby releasing the vacuole from Myo2p (step 3). Notably, this release deposits the vacuole near the centre of the bud. Moreover, the degradation of Vac17p may release Myo2p to carry out other functions including movement of secretory vesicles to the site of cytokinesis at the mother–bud neck.

Our study of Myo2p/Vac17p/Vac8p provides one example where disassembly of a transport complex ensures the proper direction of vacuole movement. Disassembly of a transport complex may also occur for kinesin-related movement where studies of KIF3a and KIF3c suggest that these motors move to the axonal terminus and do not return^{21,22}. We postulate that the regulated assembly and disassembly of transport complexes may be a common mechanism for the movement of cellular components from one location to another. □

Methods

Strains and materials

The *Saccharomyces cerevisiae* strains used for vacuole inheritance studies are derivatives of LWY7235 (ref. 23). The yeast two-hybrid assay strain was PJ69-4A²⁴. 1 × HA-MYO2 was constructed by substituting the *MluI/StuI* fragment of pNLC39 (ref. 6) with the *MluI/StuI* fragment from pN444 (yIP5-MYO2-HA)²⁵. Anti-Vac17p antibody was generated in sheep against the glutathione S-transferase (GST)–Vac17(320–425)p fusion protein and affinity purified. The rabbit anti-Vac8p (ref. 9) and goat anti-Myo2p (ref. 10) antibodies have been described previously. Mouse anti-Pgk1p, anti-Vph1p, anti-Vma2p and Alexa 488-donkey anti-goat immunoglobulin-γ (IgG) were purchased from Molecular Probes. All other antibodies were purchased from Jackson ImmunoResearch Laboratories.

Identification of VAC17 and construction of vac17 mutants

The *vac17-1* mutant (pep4[−]) containing YCP50-GAL-PEP4 was transformed with a yeast genomic DNA library (ATCC 77162). Transformants were replica-plated onto galactose medium to induce PEP4 and then transferred back to glucose plates to repress PEP4. The resulting colonies were tested for carboxypeptidase Y (CPY) activity⁶. CPY⁺ indicates normal vacuole inheritance. CPY[−] indicates a defect²⁶. 1.5 mg ml^{−1} of Fast Red Violet LB salt (Sigma) was substituted for Fast Garnet GBC. From 8,000 transformants, we obtained 20 CPY⁺ colonies (orange colonies, Fig. 1a), three of which displayed wild-type vacuole inheritance, similar to cells shown in Fig. 1b. Plasmids from these candidates contained the open reading frame (ORF) YCL063w. Sequencing the YCL063w ORF in *vac17-1* revealed that the seventh codon was changed from GAG to TAG.

pVAC17 is pRS416 with the 1.94 *HindIII* fragment containing full-length YCL063w. The *BglII/NsiI* fragment of pVAC17, which corresponds to residues 26–357 of Vac17p, was substituted with *TRP1* to generate pVT12. The *HaeIII/AvaI* fragment of pVT12 was transformed into a diploid strain (LWY 3145) to construct the *vac17Δ::TRP1* strain (LWY5798).

pVAC17(Δ97–259) was constructed by deleting the 489-base-pair (bp) *EcoRI* fragment from pVAC17. pVAC17(Δ109–190) was constructed by deleting the 243-bp *MscI/BsaAI* fragment of pVAC17. pVAC17(Δ196–251) was constructed by deleting the 168-bp *NlaIII* fragment from pVAC17. pVAC17(Δ319–377) was constructed in two steps. First, pVAC17(Δ319–355) was constructed by substituting the *BglIII/AflII* fragment of pVAC17 with the *BglIII/AflII* fragment of residues 1–318 amplified with primers TF1 (AAAAGGATCCATGGCAACCCAGCCCTAGAG) and v17C318 (CAAGCTTAAGCGCTGCAACAGTTGGATGACA). Then residues 355–377 were deleted by inserting the *AflII/PstI* fragment of residues 378–425 amplified by primers v17N378

(AAAGCTTAAGAGGGGGGAGAAAGAATAAGGGC) and TF2 (AAAAGCTGAGAAAGATGGCACCCGAGTCTAG) into the *AflII/NsiI* sites of pVAC17(Δ319–355). 3 × HA-VAC17 was constructed by inserting the 3 × HA epitope into the unique *BglII* site of VAC17 in pVAC17.

Fluorescence microscopy

To measure vacuole inheritance, cells were labelled with FM4-64 and then chased for one doubling time. Vacuole inheritance was scored as the ratio of buds with labelled vacuoles^{27–29}. Double-label immunofluorescence was performed as described²⁷ except that cells were fixed in 0.1 M phosphate buffer (pH 6.5) containing 4% formaldehyde. After formation of spheroplasts, cells were incubated with sheep anti-Vac17p IgG (1:1,000), rabbit anti-sheep IgG (1:240), goat anti-rabbit IgG (1:240), and Alexa 488-conjugated donkey anti-goat IgG (green) (1:200). The vacuole membrane was visualized with a mouse monoclonal antibody to the 60K (relative molecular mass, *M_r* 60,000) subunit (Vma2p) of the vacuolar ATPase (1:100) followed by rhodamine-red-conjugated donkey anti-mouse IgG (1:200). Images were taken with a BioRad confocal microscope⁹.

Synchronization and cell size measurement

Cells (200 ml; optical density of 0.2 at 600 nm) were synchronized by incubation with 2.5 μM α-factor (Zymo Research) for 2 h. Arrested cells were washed with fresh medium twice to remove α-factor and incubated in 400 ml of fresh medium at 24 °C. Cells (20 ml) were collected every 10 min and broken with glass beads⁹. The sample was heated at 70 °C for 5 min and stored on dry ice. All samples were reheated at 70 °C for 5 min and mixed with loading buffer, separated by 10% SDS-PAGE and transferred to a nitrocellulose membrane. Vac17p and Pgk1p were detected by anti-Vac17p antibody (1:20,000) and anti-Pgk1p (1:5,000) antibody, respectively. In addition, for each time point, images of cells viewed by brightfield microscopy were recorded by an RT SPOT camera (Diagnostic Instruments). We measured the diameters of the bud and the mother cell using the MetaMorph software (Universal Imaging).

Received 25 October 2002; accepted 28 January 2003; doi:10.1038/nature01453.

Published online 16 February 2003.

- Verhey, K. J. & Rapoport, T. A. Kinesin carries the signal. *Trends Biochem. Sci.* **26**, 545–550 (2001).
- Wells, A. L. *et al.* Myosin VI is an actin-based motor that moves backwards. *Nature* **401**, 505–508 (1999).
- Huang, J. D. *et al.* Direct interaction of microtubule- and actin-based transport motors. *Nature* **397**, 267–270 (1999).
- Govindan, B., Bowser, R. & Novick, P. The role of Myo2, a yeast class V myosin, in vesicular transport. *J. Cell Biol.* **128**, 1055–1068 (1995).
- Pruyne, D. W., Schott, D. H. & Bretscher, A. Tropomyosin-containing actin cables direct the Myo2p-dependent polarized delivery of secretory vesicles in budding yeast. *J. Cell Biol.* **143**, 1931–1945 (1998).
- Catlett, N. L., Duex, J. E., Tang, F. & Weisman, L. S. Two distinct regions in a yeast myosin-V tail domain are required for the movement of different cargoes. *J. Cell Biol.* **150**, 513–526 (2000).
- Rossanese, O. W. *et al.* A role for actin, Cdc1p, and Myo2p in the inheritance of late Golgi elements in *Saccharomyces cerevisiae*. *J. Cell Biol.* **153**, 47–62 (2001).
- Hoefner, D., van den Berg, M., Philippsen, P., Tabak, H. F. & Hettema, E. H. A role for Vps1p, actin, and the Myo2p motor in peroxisome abundance and inheritance in *Saccharomyces cerevisiae*. *J. Cell Biol.* **155**, 979–990 (2001).
- Wang, Y. X., Catlett, N. L. & Weisman, L. S. Vac8p, a vacuolar protein with armadillo repeats, functions in both vacuole inheritance and protein targeting from the cytoplasm to vacuole. *J. Cell Biol.* **140**, 1063–1074 (1998).
- Ishikawa, K. *et al.* Identification of an organelle-specific myosin-V receptor. *J. Cell Biol.* (in the press).
- Scott, S. V. *et al.* Ap13p and Vac8p are part of a complex of phosphoproteins that are required for cytoplasm to vacuole targeting. *J. Biol. Chem.* **275**, 25840–25849 (2000).
- Huber, A. H. & Weis, W. I. The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. *Cell* **105**, 391–402 (2001).
- Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
- Zhu, G. *et al.* Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature* **406**, 90–94 (2000).
- Reichsteiner, M. & Rogers, S. W. PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* **21**, 267–271 (1996).
- Gomes de Mesquita, D. S., ten Hoopen, R. & Woldringh, C. L. Vacuolar segregation to the bud of *Saccharomyces cerevisiae*: an analysis of morphology and timing in the cell cycle. *J. Gen. Microbiol.* **137**, 2447–2454 (1991).
- Lillie, S. H. & Brown, S. S. Immunofluorescence localization of the unconventional myosin, Myo2p, and the putative kinesin-related protein, Smy1p, to the same regions of polarized growth in *Saccharomyces cerevisiae*. *J. Cell Biol.* **125**, 825–842 (1994).
- Pruyne, D. & Bretscher, A. Polarization of cell growth in yeast. *J. Cell Sci.* **113**, 571–585 (2000).
- Colman-Lerner, A., Chin, T. E. & Brent, R. Yeast Cbk1 and Mob2 activate daughter-specific genetic programs to induce asymmetric cell fates. *Cell* **107**, 739–750 (2001).
- Weiss, E. L. *et al.* The *Saccharomyces cerevisiae* Mob2p–Cbk1p kinase complex promotes polarized growth and acts with the mitotic exit network to facilitate daughter cell-specific localization of Ace2p transcription factor. *J. Cell Biol.* **158**, 885–900 (2002).
- Yang, Z. & Goldstein, L. S. Characterization of the KIF3C neural kinesin-like motor from mouse. *Mol. Biol. Cell* **9**, 249–261 (1998).
- Kondo, S. *et al.* KIF3A is a new microtubule-based anterograde motor in the nerve axon. *J. Cell Biol.* **125**, 1095–1107 (1994).
- Bonangelino, C. J., Catlett, N. L. & Weisman, L. S. Vac7p, a novel vacuolar protein, is required for normal vacuole inheritance and morphology. *Mol. Cell Biol.* **17**, 6847–6858 (1997).
- James, P., Halladay, J. & Craig, E. A. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144**, 1425–1436 (1996).

25. Reck-Peterson, S. L., Novick, P. J. & Mooseker, M. S. The tail of a yeast class V myosin, myo2p, functions as a localization domain. *Mol. Biol. Cell* **10**, 1001–1017 (1999).
26. Gomes de Mesquita, D. S., van den Hazel, H. B., Bouwman, J. & Woldringh, C. L. Characterization of new vacuolar segregation mutants, isolated by screening for loss of proteinase B self-activation. *Eur. J. Cell Biol.* **71**, 237–247 (1996).
27. Hill, K. L., Catlett, N. L. & Weisman, L. S. Actin and myosin function in directed vacuole movement during cell division in *Saccharomyces cerevisiae*. *J. Cell Biol.* **135**, 1535–1549 (1996).
28. Catlett, N. L. & Weisman, L. S. The terminal tail region of a yeast myosin-V mediates its attachment to vacuole membranes and sites of polarized growth. *Proc. Natl Acad. Sci. USA* **95**, 14799–14804 (1998).
29. Catlett, N. L. & Weisman, L. S. Divide and multiply: organelle partitioning in yeast. *Curr. Opin. Cell Biol.* **12**, 509–516 (2000).

Acknowledgements We thank M. Cantwell for helping with the isolation of the *vac17-1* mutant. We thank T. Monniger and the University of Iowa Central Microscopy Research Facility for guidance in the use of the confocal microscope. We thank R. Cohen, J. Donelson, R. Piper, P. Rubenstein, J. Shaw and M. Starnes for discussions. This work was supported by grants from the National Institutes of Health and the National Science Foundation (to L.S.W.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.S.W. (e-mail: lois-weisman@uiowa.edu).

retractions

Superconductivity in CaCuO_2 as a result of field-effect doping

J. H. Schön, M. Dorget, F. C. Beuran, X. Z. Zu, E. Arushanov, C. Deville Cavellin & M. Laguës

Nature **414**, 434–436 (2001).

This manuscript was, in part, the subject of an independent investigation¹ conducted at the behest of Bell Laboratories, Lucent Technologies. The independent committee reviewed concerns related to the validity of data associated with the device measurements described in the paper.

J.H.S.: As a result of the committee's findings¹, I am issuing a retraction of the paper. I note nevertheless that this paper may also contain some legitimate ideas and contributions.

M.D., X.Z.Z., E.A. and C.D.C.: In the light of the recent findings of the investigation¹ committee chaired by Professor Beasley, we would like to warn readers about the validity of the field-effect doping data presented in this paper and issue a retraction of this article. Our laboratory specializes in the synthesis, by molecular beam epitaxy, of copper oxide thin films. In May 2001, we initiated a collaboration with J.H.S., in which our role was limited to the synthesis of a thin-film sample of CaCuO_2 . We can certify the quality (composition and structure) of the sample, and we are ready to prepare such samples for other serious scientific teams who want to try to reproduce these results.

M.L. and F.C.B.: We comment here as researchers at Wintici SA, a technology company. The synthesis of the CaCuO_2 sample reported in the paper was undertaken in collaboration with researchers from ESPCI, and we can vouch for its quality. But in the light of the committee's findings¹, we wish to issue a retraction of the paper. We note nevertheless that this paper may also contain some legitimate ideas and contributions. □

1. Beasley, M. R., Datta, S., Kogelnik, H., Kroemer, H. & Monroe, D. Report of the Investigation Committee on the Possibility of Scientific Misconduct in the Work of Hendrik Schön and Coauthors. (<http://publish.aps.org/reports/>) (doi:10.1103/aps.reports.lucent) (Lucent Technologies/American Physical Society, September 2002).

Superconductivity in single crystals of the fullerene C_{70}

J. H. Schön, Ch. Kloc, T. Siegrist, M. Steigerwald, C. Svensson & B. Batlogg

Nature **413**, 831–833 (2001).

This manuscript was, in part, the subject of an independent investigation¹ conducted at the behest of Bell Laboratories, Lucent Technologies. The independent committee reviewed concerns related to the validity of data associated with the device measurements described in the paper. As a result of the committee's findings, we are issuing a retraction of the paper. We note nevertheless that this paper may also contain some legitimate ideas and contributions. □

1. Beasley, M. R., Datta, S., Kogelnik, H., Kroemer, H. & Monroe, D. Report of the Investigation Committee on the Possibility of Scientific Misconduct in the Work of Hendrik Schön and Coauthors. (<http://publish.aps.org/reports/>) (doi:10.1103/aps.reports.lucent) (Lucent Technologies/American Physical Society, September 2002).

Self-assembled monolayer organic field-effect transistors

Jan Hendrik Schön, Hong Meng & Zhenan Bao

Nature **413**, 713–716 (2001); correction *Nature* **414**, 470 (2001).

This manuscript was, in part, the subject of an independent investigation¹ conducted at the behest of Bell Laboratories, Lucent Technologies. The independent committee reviewed concerns related to the validity of data associated with the device measurements described in the paper. As a result of the committee's findings, we are issuing a retraction of the paper. We note nevertheless that this paper may also contain some legitimate ideas and contributions. □

1. Beasley, M. R., Datta, S., Kogelnik, H., Kroemer, H. & Monroe, D. Report of the Investigation Committee on the Possibility of Scientific Misconduct in the Work of Hendrik Schön and Coauthors. (<http://publish.aps.org/reports/>) (doi:10.1103/aps.reports.lucent) (Lucent Technologies/American Physical Society, September 2002).

Gate-induced superconductivity in a solution-processed organic polymer film

J. H. Schön, A. Dodabalapur, Z. Bao, Ch. Kloc, O. Schenker & B. Batlogg

Nature **410**, 189–192 (2001).

This manuscript was, in part, the subject of an independent investigation¹ conducted at the behest of Bell Laboratories, Lucent Technologies. The independent committee reviewed concerns related to the validity of data associated with the device measurements described in the paper. As a result of the committee's findings, we are issuing a retraction of the paper. We note nevertheless that this paper may also contain some legitimate ideas and contributions. □

1. Beasley, M. R., Datta, S., Kogelnik, H., Kroemer, H. & Monroe, D. Report of the Investigation Committee on the Possibility of Scientific Misconduct in the Work of Hendrik Schön and Coauthors. (<http://publish.aps.org/reports/>) (doi:10.1103/aps.reports.lucent) (Lucent Technologies/American Physical Society, September 2002).

Superconductivity at 52 K in hole-doped C₆₀

J. H. Schön, Ch. Kloc & B. Batlogg

Nature **408**, 549–552 (2000).

This manuscript was, in part, the subject of an independent investigation¹ conducted at the behest of Bell Laboratories, Lucent Technologies. The independent committee reviewed concerns related to the validity of data associated with the device measurements described in the paper. As a result of the committee's findings, we are issuing a retraction of the paper. We note nevertheless that this paper may also contain some legitimate ideas and contributions. □

1. Beasley, M. R., Datta, S., Kogelnik, H., Kroemer, H. & Monroe, D. Report of the Investigation Committee on the Possibility of Scientific Misconduct in the Work of Hendrik Schön and Coauthors. (<http://publish.aps.org/reports/>) (doi:10.1103/aps.reports.lucent) (Lucent Technologies/American Physical Society, September 2002).

Superconductivity in molecular crystals induced by charge injection

J. H. Schön, Ch. Kloc & B. Batlogg

Nature **406**, 702–704 (2000).

Several papers were recently the subject of an independent investigation¹ conducted at the behest of Bell Laboratories, Lucent Technologies. The independent committee reviewed concerns related to the validity of data associated with the device measurements described in those papers. As a result of the committee's findings, we have issued retractions of those papers. Furthermore, because of the extensive and serious nature of the committee's findings relating to the manuscripts that they examined, we are additionally concerned

about aspects of the data presented in this paper. As we cannot vouch for the validity of the data, we wish to withdraw our support for the paper and issue a retraction.

The first author of this paper (J.H.S.), who was responsible for most of the experimental work, wishes to be dissociated from this retraction because he believes in the science presented in this manuscript. □

1. Beasley, M. R., Datta, S., Kogelnik, H., Kroemer, H. & Monroe, D. Report of the Investigation Committee on the Possibility of Scientific Misconduct in the Work of Hendrik Schön and Coauthors. (<http://publish.aps.org/reports/>) (doi:10.1103/aps.reports.lucent) (Lucent Technologies/American Physical Society, September 2002).

Efficient organic photovoltaic diodes based on doped pentacene

J. H. Schön, Ch. Kloc, E. Bucher & B. Batlogg

Nature **403**, 408–410 (2000).

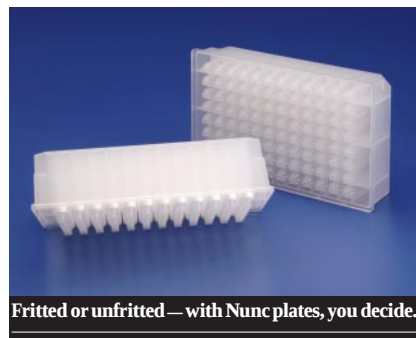
Several papers were recently the subject of an independent investigation¹ conducted at the behest of Bell Laboratories, Lucent Technologies. The independent committee reviewed concerns related to the validity of data associated with the device measurements described in those papers. As a result of the committee's findings, we have issued retractions of those papers. Furthermore, because of the extensive and serious nature of the committee's findings relating to the manuscripts that they examined, we are additionally concerned about aspects of the data presented in this paper. As we cannot vouch for the validity of the data, we wish to withdraw our support for the paper and issue a retraction.

The first author of this paper (J.H.S.), who was responsible for most of the experimental work, wishes to be dissociated from this retraction because he believes in the science presented in this manuscript. □

1. Beasley, M. R., Datta, S., Kogelnik, H., Kroemer, H. & Monroe, D. Report of the Investigation Committee on the Possibility of Scientific Misconduct in the Work of Hendrik Schön and Coauthors. (<http://publish.aps.org/reports/>) (doi:10.1103/aps.reports.lucent) (Lucent Technologies/American Physical Society, September 2002).

Cell biology in close-up

A double dose of nucleic acids, and a new adornment for the lab wall.



Filter Plates

Nunc www.nuncbrand.com

High-throughput filtering

The new filter plates in Nunc's molecular biology range are designed for high-throughput sample processing, including plasmid isolation. The 1 ml filter plates are available with or without a 20 μ m solid frit and allow the user to process samples by vacuum filtration or centrifugation, in conjunction with the Nunc 1 ml deep well plate to receive the filtrate. The fritted filter plate is suitable for the clarification of bacterial and other lysates and crude extracts. Both the fritted and unfritted filter plate can be used to develop customized purification processes by addition of supports, membranes or resins as required. The glass fibre filter plate is intended for rapid isolation of plasmid DNA from multiple samples. A step-by-step plasmid purification protocol is included in each case of the product. DNA yield depends not only on the process but also on the plasmid copy number and cell density.

CKX Series microscopes

Olympus Optical Co.

www.olympus-europa.com

Living cells under the microscope

Olympus has developed a new relief contrast (RC) technique with the low-cost CKX Series microscopes for cell observation. This has particular advantages for the observation of cell viability in unstained living specimens. The new technique uses a modulated oblique illumination to increase visibility and contrast. A modulator within the condenser combined with the phase contrast objective enhances a pseudo-relief effect and allows superb apparent three-dimensional images. These are particularly effective with plastic Petri dishes and other cell culture devices. Contrast and brightness can be adjusted by rotating the polarizer. Centering free phase contrast is provided as standard for 10 $\times$,

20 $\times$ and 40 $\times$ magnifications. It is possible to view the specimen immediately using a full set of objectives and a single slider with pre-centred phase contrast. The CKX31 is a budget microscope for cell observation, while the CKX41 allows upgrading to more advanced operations.

Cytokine wall chart

MedSystems Diagnostics

www.instant-elisa.com

Poster session

A colourful cytokine interactions wall chart is now available from Bender MedSystems Diagnostics. The chart illustrates the cytokines produced by immune cells along with their interactions in mediating inflammatory reactions — and includes a list of ELISA and Instant ELISA kits available from the company. A wide selection of Instant ELISA kits is available for detecting cytokines from human and monkey sources.

QIAamp MinElute

Qiagen

www.qiagen.com

Simultaneous purification of viral RNA and DNA

QIAamp MinElute virus spin and vacuum kits combine well-established silica-membrane technology with QIAamp MinElute columns for simultaneous purification of viral RNA and DNA. This combination allows elution in small volumes, yielding highly concentrated nucleic acids. The procedure is suitable for use with plasma, serum, and other cell-free body fluids. The procedures involved in these kits are designed to ensure that there is no sample-to-sample cross-contamination and to allow safe handling of potentially infectious samples. The simple protocols, which are suited for simultaneous processing of multiple samples, yield pure nucleic acids

in less than 1 hour. QIAamp MinElute virus spin and vacuum kits can be used for isolation of viral RNA and DNA from a broad range of RNA and DNA viruses. Nucleic acids purified using the QIAamp MinElute system are ready for immediate use in amplification reactions, such as PCR and RT-PCR, or for storage at -20°C .

These notes are compiled in the Nature office from information provided by the manufacturers.



Two for one: QIAamp MinElute.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harakatomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejp.com

naturejobs

Recruitment blueprint

Science Foundation Ireland (SFI) was established in 2000 to beef up investments in the country's sagging infrastructure for research and development. Its stated goal was to bring in new talent — but would it succeed? With a budget of E646 million (US\$698 million) to spend between 2000 and 2006, the answer seemed obvious. But the first rounds of grants attracted criticism because much of the money went to researchers who were already based in Ireland (see *Naturejobs* 4–5; 21 March 2002).

A few years on, there are signs that the scheme may be paying off. One of the more promising results is the SFI-supported nanotechnology laboratory that opened at Dublin's Trinity College in January. This lab will receive E33 million from the SFI over five years, but more important than the money is the lab's composition.

John Pethica was recruited from the materials department at the University of Oxford, Suzi Jarvis relocated from the National Institute of Advanced Industrial Science and Technology in Japan, and John Boland moved from the University of North Carolina at Chapel Hill. Those three, along with Michael Coey and Igor Shvets, who were already at Trinity before the SFI money came through, will lead a strongly international lab. The multidisciplinary team consists of 54 people from 19 countries on four continents.

That kind of recruitment is good news for other Irish labs seeking to follow suit, as Ireland is a country that has only relatively recently begun serious investment in R&D and doesn't have much of a tradition of recruiting internationally. The news must also be encouraging for other countries that hope they can fill their sparkling new facilities with researchers from abroad.

Paul Smaglik

Naturejobs editor



Contents

POSTDOCS AND STUDENTS

Visa trouble in the
United States p96

MOVERS

Shuffles at the National
Institutes of Health; personnel
boost at the Stowers Institute;
stem-cell research heads west;
and more p98

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

Access delayed

Tight new security measures aimed at preventing terrorism mean that postdocs and students travelling to the United States need to make plans a long way in advance and expect delays, says Karen Kreeger.



Recent revisions to a website hosted by the US National Academies are a clear sign of the times for scientists wishing to enter the United States. The site provides information on how to get a visa, and its advice on what to do in the event of a visa delay or refusal is increasingly important. Heightened security measures mean that applications can now take several months to review.

Some people need hardly bother — a web link to the US Department of State warns of stringent checks for “aliens from state sponsors of terrorism” — defined as North Korea, Cuba, Syria, Sudan, Iran, Iraq and Libya. “Last May, Congress passed a law stating that every person who was from one of the countries on a list of state sponsors of terrorism must be put through a very thorough review,” says Stuart Patt, a spokesman for consular affairs at the state department.

But many scientists from other countries have been unable to take up posts or attend conferences since the rules changed last summer. And some who study in the United States have found that they cannot return after a visit abroad.

Waiting for a visa is an old problem, says Patt. But it

has been exacerbated since 11 September 2001. “There is a new reality for scientists and students who require visas,” says Wendy White, director of the National Academies’ Board on International Scientific Organizations. “Understanding all they can about the systems before applying will be their best defence, and allowing at least three months for the process to work.”

More visas are being reviewed for security reasons, White explains. If there are any doubts by the embassy issuing the visa, the application is sent to Washington for checking. “It used to be that, if there were no explicit notifications from the FBI or CIA within a certain period of time, the visa could be granted,” says White. “It was like an implicit approval.” But now there must be explicit approval from such federal agencies as the FBI and CIA in some cases, and an application might need approval by more than one agency. “That’s where this legendary backlog comes in,” says White.

BATTLING THE BACKLOG

One component of the backlogs, according to John Marburger, director of the White House Office of Science and Technology Policy, is that consular officials in the visa holders’ country of origin are personally liable for the actions of the people that they let into the United States. That law was passed after the first World Trade Center bombing in the early 1990s. “Now it’s being taken very seriously and it’s a contributing factor,” he says.

Marburger says that the administration is “concerned” about the backlog, but adds that the security measures are necessary. “We don’t want to stop or diminish the international flavour of our research, but the public does expect us to have reasonable measures to protect the populace against terrorism,” he says. To help speed up visa applications, Marburger’s office is convening working groups to identify bottlenecks and provide better information for both visa applicants and processors. But he does not expect the backlog to be cleared within the next few weeks.

Foreign nationals need to allow more time to obtain US visas, says Wendy White.



Meanwhile, some students and postdocs have had their research pushed back because they had problems returning to the United States. Anna Marie Skalka, scientific director of basic science at Fox Chase Cancer Center in Philadelphia, has a Russian student who was stuck in Moscow for three months. The centre has a partnership with the Russian State Medical University, in which students do research at Fox Chase and their degree is conferred by the Russian institute. The student travelled to Russia to defend his thesis, with plans to return to Philadelphia on 21 December. But his visa for re-entry was delayed until February. Even though he has made it back, the student is anxious about completing his work. His 'J1' exchange visitor visa — allowing temporary entry for a specific educational objective — runs out in May.

"It's just bewildering," says Skalka. "We don't do any classified research. This is a cancer centre; we're working on basic science."

Last spring, her student returned to Russia for his preliminary defence and returned with no problems. "What's happened is that a lot of these young people went home to their families for Christmas and they haven't been able to come back," says Skalka. "I can understand that there's a need for security, but this is hurting ourselves. It's the law of unintended consequence." She hopes that some group will come forth on a national level to point out that this is a problem for the whole research enterprise.

Indeed, the presidents of three national academies issued a statement in December 2002 urging the US government not to obstruct entry by the sort of people whose work has traditionally benefited the United States.

"The US scientific, engineering, and health communities cannot hope to maintain their present position of international leadership if they become isolated from the rest of the world," wrote Bruce Alberts of the US National Academy of Sciences, William Wulf of the National Academy of Engineering, and Harvey Fineberg of the Institute of Medicine. They noted that, as well as students and postdocs, those who couldn't



Anna Marie Skalka (second left, front row) has a number of foreign nationals in her research team, but she is finding that new visa restrictions can cause problems.

enter the country included experts invited to teach or speak at conferences, and even foreign associates of the academies (see 'Out in the cold', below).

"I think there's an overall feeling of insecurity," says Sharon Ladd, director of the International Office at Harvard University in Cambridge, Massachusetts. Ladd says that at Harvard it has been mostly graduate students who have been held up. She cites as an example a Sudanese law student who, while working in Europe last summer, went to London to visit friends and family. He was finally able to return to Harvard in mid-January after being stalled for five months. "But I couldn't point a finger and say this particular lab has this number of people affected in this way," she says.

"We agree that some things need to be done," says White. "The security concerns are out there and they aren't going to go away." But ultimately, she adds, "we're worried about the chilling effects on science and the unintended consequences and actions that are put in place in the name of security."

Karen Kreeger is a freelance science writer based in Philadelphia.

Web Links

National Academies' visa information site

♦ www7.nationalacademies.org/visas

National Academies presidents' statement on visa restrictions

♦ www4.nationalacademies.org/news.nsf/isbn/s12132002?OpenDocument

Out in the cold

The new visa checks are affecting more than just postdocs and graduate students. Many distinguished scientists are hitting the same roadblocks when trying to attend meetings in the United States.

Caroline Herzenberg, a retired scientist from Argonne National Laboratory in Illinois, relates the story of Valery Nesvizhevsky of the Laue-Langevin Institute in Grenoble, France. Nesvizhevsky works on quantum states of ultracold neutrons in Earth's gravitational field. "He was invited to speak at Argonne National Laboratory in early to mid-2002, and was denied a visa; he was invited again for an 8 November 2002 colloquium, and again denied a visa," says Herzenberg. It was only when he was invited for a third



Caroline Herzenberg: worried by red tape.

important experiments, from coming to talk about his work to other physicists here in the United States has had, in its own way, a negative effect on research in this country," Herzenberg says.

Visa restrictions also disrupted the annual symposium of Chinese-American

time to speak both at Argonne and the University of Chicago in mid-January 2003 that he was allowed to come.

"I would say that the US government repeatedly preventing this physicist, who has conducted such fundamental and

Frontiers of Science, a joint programme between the US National Academies and the Chinese Academy of Sciences, which was scheduled for 11–13 October last year in Irvine, California.

The Chinese delegation applied for its visas in early August, but by mid-September still had not received them. After officials contacted the US state department, the visas were issued on 9 October, but too late for the group to travel the next day. The meeting was postponed until late November, which meant that eight of the original 24 presenters could not take part. Even with four replacement presenters, attendance was reduced from 77 to 50 and the delay cost the National Academies nearly \$40,000 in travel and hotel expenses.

K.K.

Transitions

X Steve Holmes has been appointed as senior director, therapeutic targets at Domantis, a drug-discovery company based in Cambridge, UK. Holmes was formerly director of antibody discovery at Oxford GlycoSciences.

X Oxford drug-discovery firm Etiologics has appointed **David Campbell** as chief scientific officer. He was previously director and head of molecular genetics, Europe, at GlaxoSmithKline.

X Natch Pharmaceutical in Bothell, Washington, recently appointed **Jade Brown** as senior director, marketing and business development.

X Reinhard Schneider, a founder and chief information officer of LION Bioscience in Heidelberg, Germany, is leaving the company.

X Astex Technology, a drug-discovery firm in Cambridge, UK, has appointed **David Rees** director of medicinal chemistry.

X Bernard Silverman, professor and provost of the Institute for Advanced Studies at the University of Bristol, UK, and a member of the government's GM Science Review Panel, will become master of St Peter's College, Oxford, in October.

X Jim Kennedy, lecturer at the department of Earth sciences and curator of the University of Oxford's geological collection, is the new director of the Oxford University Museum of Natural History. He succeeds **Keith Thomas** in October.

X Mark Rohrbaugh has been appointed director of the Office of Technology Transfer in the Office of Intramural Research at the US National Institutes of Health.

MEDICAL RESEARCH

Last month, **Raynard Kington** was appointed as deputy director of the US National Institutes of Health (NIH). He replaces **Ruth Kirschstein**, who had been in the post since 1993. Kirschstein, who was also acting director for the NIH from January 2000 to May 2002, will now be the senior adviser to the agency's director, **Elias Zerhouni**.

Kington had been associate director for behavioural and social-sciences research at the NIH, as well as director of the NIH Office of Behavioral and Social Sciences Research since November 2000. He also served as acting director for the National Institute on Alcohol Abuse and Alcoholism for nine months in 2002.

He came to the NIH from the Centers for Disease Control and Prevention, where he oversaw surveys of the health and related behaviour of the US population.

After 30 years at the NIH, **Wendy Baldwin** has returned to her Alma Mater. Baldwin has been deputy director for extramural research at the NIH since 1993, but this January she joined the University of Kentucky as vice-president of research. There she will oversee an enterprise that increased its grants and contracts awards by 22% for the year 2001–02. Such growth is not unfamiliar to Baldwin — she oversaw extramural grants at the NIH during a period of rapid growth for the agency's budget.

The US National Institute on Drug Abuse in Bethesda, Maryland, has a new director. Starting next month, **Nora Volkow** will take over the reins from **Glen Hanson**, who has been acting director of the institute since December 2001.

Volkow received her medical training in Mexico, but came to the United States to do a postdoc in psychiatry at New York University and has stayed in the state ever since. She is currently associate director for life sciences at the Brookhaven National Laboratory and associate dean for the medical school at the Stony Brook University in New York.

STEM-CELL RESEARCH

Evan Snyder, who is perhaps best known for his work in neural migration and plasticity, last month joined the Burnham Institute in La Jolla, California, as director of its stem-cell and regeneration programme.

Snyder is making the move from Harvard Medical School where, in his words, he has spent "half my life and my entire career". Although he stresses that he has enjoyed his time at Harvard, Snyder admits that he had become increasingly



Raynard Kington



Nora Volkow



Susan Abmayr



Michael Washburn



Jerry Workman

aware that a lot of work in his field was progressing at "nimble" research institutions, rather than at universities. Many of these institutes, like Burnham, were in the San Diego area, he adds.

California has an additional attraction for Snyder — the state government has passed a law saying that stem-cell research should be unfettered by politics. Although this legislation is largely symbolic, it signalled that the state is unlikely to interfere with his plans for stem-cell research at Burnham. These plans include the possible development of facilities to derive stem cells from human embryos — work that can only be funded by private money according to current US law.

GENETICS AND PROTEOMICS

The Stowers Institute for Medical Research in Kansas City, Missouri, has added another three scientists to its roster, bringing its total headcount to 18. The centre's good facilities, sizeable endowment (about \$1 billion) and lack of tenure are proving to be a major draw.

Susan Abmayr, who joins the institute in July as an associate scientist, was attracted by the prospect of easy access to equipment. At Pennsylvania State University in University Park, where she is currently associate professor of molecular genetics, faculty members have to share a confocal microscope, Abmayr notes. At Stowers, she says, she will have her own imaging equipment.

An added attraction is the institute's focus on science, which Abmayr feels will allow her to pursue her work on the genetics of muscle development unimpeded by other duties. "I like teaching," she says. "But between teaching and committee work, you spend so much time away from the science in your lab."

Michael Washburn will also join the institute in July, when he will leave the Torrey Mesa Research Institute in San Diego to become Stowers' director of proteomics. He will be responsible for setting up and running the institute's proteomics facility.

He says that he was attracted to the position because it will allow him to manage a facility as well as conduct his own research. He is also enthusiastic about the rapidly growing reputation of the institute, which only opened its doors in 2000.

Jerry Workman, who has been at Pennsylvania State University since 1992, is the third of the institute's new recruits. He comes in as a senior scientist. Workman is a leader in studying how chromatin, the protein-associated form of DNA, modulates gene expression.

CONTACTING US AT MOVERS

Please send any information on appointments or job moves to:
 ◆ naturejobseditor@naturedc.com